Birthday Paradox

Marek Ulm

August 2023

1. Introduction to the Illusion

Have you ever pondered the likelihood of encountering a pair who share the same birthday in a group? Picture this: a casual gathering of friends, just 23 strong. Astonishingly, there's roughly a 50-50 chance that within this group, a pair of individuals will surprisingly share the same birthday.

Welcome to the intriguing world of the birthday paradox, where probability often defies our intuition. And now, you have the chance to explore this phenomenon firsthand using the interactive plot above.

Source: Kagan (2013)

2. History

The English mathematician, Harold Davenport, also known for his expertise in number theory, is generally attributed to the problem, although he didn't publish it at that time. Davenport didn't assert himself as its discoverer because he couldn't imagine that it hadn't been stated earlier.

Source: Singmaster (2004)

The initial publication of a version of the birthday problem was by Richard von Mises in 1939, an Austrian mathematician, famous for his work in fields like statistics, probability theory, and many more.

Source: von Mises (1939), *Birthday Problem* (2023)

3. Mathematical Modeling

To comprehend and mathematically model the problem, we need to establish the following assumptions:

3.1. Simplification of year and day:

To facilitate our analysis and delve into the intricacies of the problem, we will also overlook the specific year itself. Our focus will solely be on the day within the year, effectively disregarding leap years and calendar fluctuations.

We'll overlook the existence of leap days, leading us to treat the year as having a fixed 365 days.

3.2. Uniform distribution of birthdays:

In our exploration, we assume that all birth dates are equally probable and unrelated to each other. This implies that the occurrence of one person's birthday doesn't influence the likelihood of another person's birthday falling on a specific day.

In essence, we consider the situation as if there are no twins or other significant factors impacting the distribution.

Just as a comment: In real-life birthdays are not equally distributed, for example there are more births in summer than in winter. As a result this clustering would make the probability of coincidence even more likely.

Source:Mathe-Online (2023)

4. Mathematical Derivation

4.1. Probability of a birthday match:

If $\mathbf{P(A)}$ represents the probability that at least two individuals in a group were born on the same day, where the event $\mathbf{A}$ is the probability of finding a group of 23 people with at least two people sharing same birthday. However it is more straightforward to compute using $\mathbf{P(\overline{A})}$,
the complementary probability that all 23 individuals in the room were born on different days. The relationship is as follows:

$\mathbf{P(A) = 1 - P(\overline{A})}$.

For the purpose of the following discussion, we number the 23 persons from 1 to 23. All of them having a different birthday, boils down to 22 constraints (the first person is unconstrained); the second person can not have the same birthday as the first one, the third person, can not have the same birthday as the first and second one and so on. Let us name these events $\mathbf{P(1), P(2)}$ up to $\mathbf{P(23)}$.

For $\mathbf{P(1)}$, no birthday is yet occupied by another person, as a result the probability for no coincidence is $\mathbf{\frac{365}{365}}$ or 1.

For the second person, $\mathbf{P(2)}$, there are fewer possibilities: they must have been born on one of the other 364 days. Hence, the probability for $\mathbf{P(2) = \frac{364}{365}}$.

This pattern continues for $\mathbf{P(3)}$ and the remaining individuals.

In our example with 23 people:

$$\mathbf{P(\overline{A})} = P(1) \cdot P(2) \cdot \ldots \cdot P(23) = \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{343}{365} \approx 0.493$$

$$\Rightarrow \quad \mathbf{P(A)} = 1 - P(\overline{A}) \approx 1 - 0.493 \approx 0.507$$

With N people, this leads to the following formula:

$$\mathbf{P(A)} = 1 - P(\overline{A}) = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{365 - N + 1}{365}$$

Some transformations yield:

$$\mathbf{P(A)} = 1 - P(\overline{A}) = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{365 - N + 1}{365}$$

$$\Leftrightarrow$$

$$\mathbf{P(A)} = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365 - (N - 1))}{365^N}$$

$$\Leftrightarrow$$

$$\mathbf{P(A)} = 1 - \prod_{k=1}^{N-1} \left(1 - \frac{k}{365}\right)$$

In the second question, you guessed the probability of a match for $\mathbf{N} = 40$, which would be almost at an astonishing 90% !

Source: *Birthday Problem* (2023), MatheGuru (2011)

5. Extensions

5.1. Generalized formula:

Beyond its application to the birthday problem, this formula (see also 4.) holds the potential to extend its utility to model various other instances of coincidences.

For instance, by introducing the variable **c** as a generalized value, where **c** represents the number of categories, in place of the constant 365, the formula's versatility expands to encompass a wider range of scenarios.

This approach allows the formula to adapt and be applied to different contexts, enabling insightful analysis across diverse situations.

$$\mathbf{P}(\mathbf{A}) = 1 - \prod_{k=1}^{N-1} \left(1 - \frac{k}{c}\right)$$

An illustrative scenario could involve a group of N individuals, where two of them coincidentally share the same favorite element from the periodic table. In this case, one might consider setting c equal to 118, representing the total number of elements in the periodic table.

Source: Diaconis & Mosteller (1989)

This also has been incorporated into the interactive plot provided above (light-grey/ upper distribution and the sliders on the left hand side of the probability distribution diagram)

5.2. Approximation formula for a match:

While this formula lends itself to straightforward calculations, its conceptual implications can be intricate to grasp. Particularly, when **c** is considerably large and **N** remains modest compared to $\mathbf{c}^{\frac{2}{3}}$, the following approximation becomes valuable. In such scenarios, the likelihood of a match can be approximated as follows:

$$\mathbf{P}(\mathbf{A}) \approx 1 - e^{\left(\frac{-(N)^2}{2*c}\right)}$$

Source: Diaconis & Mosteller (1989)

5.3. Probability share same birthday as you:

In the context of the birthday problem, it's important to note that neither of the two individuals is preselected in advance. On the other hand, if we shift our focus to a scenario where we specifically consider the probability $\mathbf{q}(\mathbf{N})$ that someone within a group of **N** other individuals shares the same birthday as a particular person (for instance, yourself), this probability is expressed as follows:

$$\mathbf{q}(\mathbf{N}) = 1 - \left(\frac{365 - 1}{365}\right)^N$$

and generalized with **c**:

$$\mathbf{q}(\mathbf{N}) = 1 - \left(\frac{c - 1}{c}\right)^N$$

Considering the conventional scenario with $\mathbf{c} = 365$, when $\mathbf{N} = 23$ is substituted into the formula, the calculated probability is approximately 6.1%.

Source: *Birthday Problem* (2023)

In the first question you guessed with values: **c** = 365, **N** = 70 that the probability is still less then 17.5%.

For your illustration, this question has also been implemented, allowing you to utilize it for testing and experimenting with various values. (dark-grey/ lower distribution and sliders on the left hand side of the probability distribution diagram)

6. The Birthday Problem Paradox: Challenging Intuition

The Birthday Problem presents a paradox that defies our initial perceptions of probability. While it might seem improbable for two people to share a birthday in a small group, the reality is quite different.

The paradox emerges from the exponential growth of potential matches as the group size increases. This exponential rise creates a counterintuitive outcome: even in modest groups, the chances of a shared birthday are remarkably high, they are assymptotically approaching one.

The paradox is a reminder that our intuitive understanding of probability can lead us astray. It reveals the disparity between our assumptions and mathematical realities. In essence, the Birthday Problem teaches us that seemingly straightforward questions can yield unexpected and paradoxical results, urging us to approach probability with a more critical perspective.

7. Conclusion: Unveiling Probability's Intricacies

The Birthday Problem stands as a captivating example of how probability can defy our intuition. It exposes the subtle complexities that underlie seemingly simple questions. This paradox reminds us that our initial assumptions about likelihoods can be misleading, and that exploring the mathematical underpinnings of everyday scenarios can lead to astonishing revelations.

Ultimately, the Birthday Problem serves as an invitation to delve deeper into the world of probability, where surprises and paradoxes await those who dare to question their intuitions.

Source: Ross (2014)

8. References

Birthday problem. (2023). `https://en.wikipedia.org/wiki/Birthday_problem`.

Diaconis, P. & Mosteller, F. (1989). Methods for studying coincidences. *Journal of the American Statistical Association*, *84* (408), 853–861. `https://www.stat.berkeley.edu/~aldous/157/Papers/diaconis_mosteller.pdf`.

Kagan, D. (2013, July). Bring science home: Probability—the birthday paradox. *Scientific American*, *309* (1), 68. `https://www.scientificamerican.com/article/bring-science-home-probability-birthday-paradox/`.

MatheGuru. (2011). *Birthday problem.* `https://matheguru.com/stochastik/geburtstagsproblem.html`.

Mathe-Online. (2023). *The birthday problem.* `https://mathe-online.at/mathint/wstat1/i_geburtstagsproblem.html`.

Ross, S. M. (2014). *A first course in probability.* Pearson.

Singmaster, D. (2004). Sources in recreational mathematics: An annotated bibliography. *Eighth Preliminary Edition*. (Section 8.B)

von Mises, R. (1939). Probability, statistics and truth. *Psychometrika*, *4* (4), 196–211.