

The Birthday Paradox

Chloe Swadling

March 2019

Abstract

This paper teaches and expands on the basics of probability using the birthday paradox. This was chosen as it is an interesting topic with surprising results, which should encourage beginners to read on. The paper shows that if there are 23 people in a room, then there is a 0.5 probability that people in the room share the same birthday. A general formula is then given which can be applied for different desired probabilities, numbers of items and numbers of possible outcomes.

Introduction

It is easy to see why most of the population doesn't understand mathematics, or even dislikes it. On many occasions, the answer that most people would find logical is nowhere close to the actual answer. One example of this is the birthday paradox. This paper will explain this paradox so that anyone with limited mathematical knowledge can understand it.

So, what actually is the birthday paradox? It is the idea that even in a small room full of people, it is highly likely that at least two people will share the same birthday.

Is it logical?

Why is it that the birthday paradox seems so irrational? It is mainly down to the fact that mankind is simple-minded and overall are self-centred. When in a room of 23 people, most of us will simply consider our own birthdays against everyone else's. This leads to only 22 comparisons. When making only 22 comparisons, it is obvious why you would most likely presume that everyone has a different birthday to you. In fact, with only 22 comparisons, there is approximately a 0.06 probability that there would be a matching birthday. This isn't all we need to consider though. We also need to look at the comparisons between everybody else. By only considering your own comparisons, you are actually completely ignoring an extra 231 pairings. This quite quickly shows us why this paradox is more logical than first thought.

Flipping a coin example

To illustrate how probabilities work, we are going to look at the simple example of flipping a coin [5]. Let's talk about the likelihood of landing on a head.

$$P(\text{getting a head}) = \frac{\text{Amount of heads on a coin}}{\text{Amount of heads and tails on a coin}} = \frac{1}{2} = 0.5 \quad (1)$$

The following formula shows the probability of any event occurring:

$$P(\text{any event}) = \frac{\text{Number of ways it can happen}}{\text{Total number of outcomes}} \quad (2)$$

Due to the fact that calculating a basic probability is done by division, it is a common mistake that people presume you can continue using division to calculate the probability that an event will occur multiple times. Most people could tell you that the chance of getting one head is 0.5 and getting two heads in a row is 0.25. Does this mean that as it is half as probable that you will throw two heads in a row as one, that you are ten times less likely to throw ten heads in a row? Anybody with some knowledge of probability will tell you that this is obviously not the case. Each time you flip a coin, the probability of landing on a head in all throws is exponentially less. This means for independent events, to calculate the probability of two events happening, you must multiply the probability of each event together.

$$P(H) = (0.5)^1 = 0.5 = \frac{0.5}{1} \quad (3)$$

$$P(H \cap H) = P(H)^2 = (0.5)^2 = 0.25 = \frac{0.5}{2} \quad (4)$$

$$P(H \cap H \cap H) = P(H)^3 = (0.5)^3 = 0.125 \neq \frac{0.5}{3} \quad (5)$$

$P(H \cap H)$ is the notation for calculating the probability of throwing two heads. These calculations show how it might be easy to presume that probabilities can be simply calculated by division but this is not accurate. You might think that this is all quite irrelevant but it is important to show how probabilities work and how we need to use exponents to calculate the likelihood of matching birthdays. An exponent tells you how many times a number is used in a multiplication [4]. The equation below shows how an exponent works. The number 2 tells us that 0.5 is used twice in the equation.

$$(0.5)^2 = 0.5 \cdot 0.5 \quad (6)$$

Explanation of the Birthday Paradox

In a group of 23 people, we will have 253 pairs to look at. A pair is a matching of two people in the room. Each pair will be checked individually to see if they have matching birthdays.

The first person has 22 comparisons to make, as they cannot be compared with themselves. This is then reduced by one for the second person as they have already been compared to the first. The comparisons continue to decrease by one until everyone has been compared. This shows that the amount of combinations is the sum of 1 to 22. This can be simply calculated as:

$$1 + 2 + \dots + n = \frac{n(n+1)}{2} \quad (7)$$

Therefore, we get 253 pairs using the following calculation:

$$\frac{22 \cdot (22+1)}{2} = 253 \quad (8)$$

The probability that two people have different birthdays is:

$$1 - \frac{1}{365} = \frac{364}{365} = 0.997 \quad (9)$$

This is down to the fact that probabilities must always be equal to one and there is only one birthday out of three hundred and sixty-five that has already been accounted for.

So, what is the probability of no matching birthdays in 253 combinations? This can be calculated using the following:

$$\left(\frac{364}{365}\right)^{253} = 0.4995 \quad (10)$$

As we just said, probabilities will always total to one, so we can say that the probability of finding a match is $1 - 0.4995 = 0.5005$. This gives roughly a probability of 0.5 of finding a match. For clarity, a 0.5 probability is the same as saying that there is a 50% chance.

In general, the probability of finding a match for any number n , where n is the number of people in the sample minus 1, is:

$$P(n) = 1 - \left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}} \quad (11)$$

Independence

The next section of this paper is a little more complex. All of the calculations shown above are assuming that birthdays are independent of each other. All that we have checked is that each pairing doesn't have a matching birthday. But is it possible for more than two people to have the same birthday? Obviously, yes. Therefore, we cannot just consider pairings and we must work out the probability of every single person within the room having a unique birthday. This isn't too important when the number of people is small compared to the sample size, as multiple matches are less likely. Although the chances of this occurring are remote, it can happen so let's look at the actual numbers involved. The below figure shows the formulas for the real probabilities.

Figure 1

Person	Chance of having a unique birthday
First	1
Second	$1 - \frac{1}{365}$
Third	$1 - \frac{2}{365}$
Twenty-Third	$1 - \frac{22}{365}$

In order to find the real probabilities, we can use the following formula:

$$p(\text{unique}) = 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right) \quad (12)$$

The above formula tells us that we have a 0.493 probability of having 23 unique birthdays. This therefore leads us to a 0.507 probability of having a match compared to the originally calculated 0.5005. As you can see, as the number of people in the room was small, the original calculation was pretty accurate.

The above $p(\text{unique})$ formula can be long-winded to calculate so we can use an approximation in order to simplify this. For this, we will use Taylor approximation. Taylor approximation says that when x is close to 0,

$$e^x \approx 1 + x \quad (13)$$

This is the same as saying:

$$e^{-x} \approx 1 - x \quad (14)$$

This is equivalent to saying that e^x is almost the same as $1+x$.

This is helpful as we can say that

$$1 - \frac{1}{365} \approx e^{-\frac{1}{365}} \quad (15)$$

This enables us to write our formula as:

$$p(\text{unique}) \approx 1 \cdot e^{-\frac{1}{365}} \cdot e^{-\frac{2}{365}} \cdots e^{-\frac{n}{365}} \quad (16)$$

It is also a rule with exponentials that:

$$e^a \cdot e^b = e^{a+b} \quad (17)$$

This now means we have:

$$p(\text{unique}) \approx e^{-\frac{(1+2+\cdots+n)}{365}} \quad (18)$$

Even simpler than this, we have already stated that:

$$1 + 2 + \cdots n = \frac{n(n+1)}{2} \quad (19)$$

Therefore, we can say that:

$$p(\text{unique}) \approx e^{-\frac{(22 \cdot 23)}{2 \cdot 365}} = 0.499998 \quad (20)$$

We can see that the approximation is very close to what we have previously calculated and we can, therefore, say that it is accurate.

General Formula

We can now give a general formula so that you can use what you've learned for n people and x total unique options available.

$$p(\text{unique}) \approx e^{-\frac{(n^2)}{2x}} \quad (21)$$

As you can see, we are using n squared rather than the originally given n times n+1. This is due to the fact that we have been approximating throughout so we don't need to use the exact figure and this is a close enough approximation. The lower the n value, the closer the approximation will be. Also, using n squared removes the multiple levels of n which allows us to rearrange the formula to give us a final n figure. Using this formula, we can approximate an n value that will give a 0.5 probability of a match.

$$p(\text{unique}) \approx e^{\frac{-(n^2)}{2x}} \quad (22)$$

$$1 - p(\text{match}) \approx e^{\frac{-(n^2)}{2x}} \quad (23)$$

$$1 - 0.5 \approx e^{\frac{-(n^2)}{2x}} \quad (24)$$

$$\ln(0.5) \approx \frac{-(n^2)}{2x} \quad (25)$$

$$-2x \ln(0.5) \approx n^2 \quad (26)$$

$$n \approx 1.177\sqrt{x} \quad (27)$$

$$n \approx 1.177\sqrt{365} \approx 22.5 \quad (28)$$

This formula can now be used to approximate many different scenarios. Another example we can use here is roulette. There are 37 different outcomes on a roulette wheel so if we take $x=37$, we get $n=9$. Therefore, in 9 spins of the wheel, you have a 0.5 probability of the ball stopping on the same number multiple times. You can also expand this formula further so that you can change the desired chance of a match. At the moment, this formula is to find the n that gives a 0.5 probability of a match. The below formula can be used for any chosen probability.

Let m equal the desired chance of a match, then:

$$n \approx \sqrt{-2 \ln(1 - m)} \cdot 1.177\sqrt{x} \quad (29)$$

We can use birth months to show how this formula works. Say we want to work out how many people we need to get a 0.75 probability of having a matching birth month. This gives us the following formula:

$$n \approx \sqrt{-2 \ln(1 - 0.75)} \cdot 1.177\sqrt{12} \approx 6.8 \quad (30)$$

This shows us that we need approximately seven people in order to have a 0.75 probability of having at least two individuals born within the same month.

Real World Observation

This paper will now look into whether this paradox genuinely occurs in real life. As previously shown, it is thought that in a group of 23 people there is a fifty percent chance of having a matching birthday within the group. Which popular groups of people contain 23 members? A football squad. In order to show real-life cases of this paradox, the paper will look into the 2018 FIFA World Cup squads [3]. There were 32 teams in the 2018 FIFA World Cup so we would, therefore, expect about 16 teams to have shared birthdays. It was found

that in the 2018 FIFA World Cup, 15 out of the 32 teams had shared birthdays within them. This is 46.9%, which is close to what we were expecting to see and helps to show that this paradox is actually accurate. Interestingly, in some teams it was found that there was actually more than one set of matching birthdays. This just helps to show that it is actually very common to find matching birthdays within small groups of people. The number of sets of matching birthdays for each team is shown in Figure 2.

Figure 2

Team	Number of sets of matched birthdays
Australia	1
Argentina	0
Belgium	0
Brazil	2
Columbia	0
Costa Rica	1
Croatia	1
Denmark	0
Egypt	0
England	1
France	1
Germany	1
Iceland	0
Iran	1
Japan	0
Mexico	0
Morocco	1
Nigeria	1
Panama	0
Peru	0
Poland	4
Portugal	3
Russia	1
Saudi Arabia	0
Senegal	0
Serbia	0
South Korea	1
Spain	1
Sweden	0
Switzerland	0
Tunisia	0
Uruguay	0

Limitations

There are limitations for these formulae. The first issue is that it is presumed that there are 365 days in a year. Obviously, we know that on a leap year there are 366 days. We can change the x value to 366 but this would suggest that there is an equal probability of being born on the 29th February as there is any other day. This is obviously false.

It is also assumed that the probability of being born on each day of the year is the same. Although this seems to be correct, there are often peaks and troughs when it comes to births. There can also be “baby booms” after large scale events, (often big sporting events), that will affect the percentage of births each day. An example of this is FC Barcelona’s dramatic last-minute win in the semi-finals of the UEFA Champions League 2009. Nine months after this event, the birth rates in Barcelona had increased by 45% [6]. This shows how the average birth rate on each individual day is unlikely to be consistent.

Looking at data into the skew of birth dates, you see that a larger number of the most common birthdays are in September [8]. This is easily explained. These dates are 9-months after the Christmas period. These are periods that are deemed to be romantic and are also times of the year where people are likely to have time off of work. Also, in some cases, people chose to not conceive so that they do not give birth on certain days of the year. It is shown that Christmas Eve and Christmas Day are two of the lowest days in relation to the average birth rates. All of these show how it is not correct to assume that the chance of being born on each day has the same probability.

An easy way of explaining this limitation is through the exaggerated model of football fans. Using the general formula, you can use $x=20$ for the 20 teams that play in the Premier League. This gives an n value (previously shown as the number of people in a room when using the general formula), of just over 5. We can, therefore, say that if 6 football fans are in a room then you have at least a probability of 0.5 of two people supporting the same team. This doesn’t work, as the number of supporters of each Premier League team is massively different and therefore the probabilities of a fan supporting each team varies greatly. According to statistics, Manchester United FC have around 73 million supporters compared to Huddersfield Town A.F.C, who have about 140 thousand [7]. You can easily see from these numbers that the probabilities of finding a supporter from each club are greatly different. Another thing to note is that with some examples the location of a study will have an impact. Quite obviously if the above study was undertaken in Liverpool, then the data will be skewed towards Liverpool fans. This doesn’t apply when studying birthdays but is something to think about when using the general formula for other problems.

Conclusion

This paper introduces basic concepts in probability by explaining the Birthday Paradox. It is seen that you have a probability of 0.5 of sharing a birthday when comparing 23 random

people. We then showed how the formula can be applied in different situations with varying probabilities. We also see how many questions can be approximated accurately allowing us to use much simpler calculations. This helps us to show how mathematics can often be much easier than first thought.

References

- [1] Betterexplained.com. (2019). *Understanding the Birthday Paradox – BetterExplained*. [online] Available at: <https://betterexplained.com/articles/understanding-the-birthday-paradox/> [Accessed 18 Mar. 2019].
- [2] Buddies, S. (2019). *Probability and the Birthday Paradox*. [online] Scientific American. Available at: <https://www.scientificamerican.com/article/bring-science-home-probability-birthday-paradox/> [Accessed 18 Mar. 2019].
- [3] En.wikipedia.org. (2019). *2018 FIFA World Cup squads*. [online] Available at: https://en.wikipedia.org/wiki/2018_FIFA_World_Cup_squads [Accessed 18 Mar. 2019].
- [4] Mathsisfun.com. (2019). *Definition of Exponent*. [online] Available at: <https://www.mathsisfun.com/definitions/exponent.html> [Accessed 22 Mar. 2019].
- [5] Mathsisfun.com. (2019). *Probability*. [online] Available at: <https://www.mathsisfun.com/data/probability.html> [Accessed 19 Mar. 2019].
- [6] Montesinos, J., Cortes, J., Arnau, A., Sanchez, J., Elmore, M., Macia, N., Gonzalez, J., Santisteve, R., Cobo, E. and Bosch, J. (2013). Barcelona baby boom: does sporting success affect birth rate?. *BMJ*, 347(dec17 9), pp.f7387-f7387.
- [7] Stadium-maps.com. (2019). *Ranking of English Premier League teams popularity*. [online] Available at: http://www.stadium-maps.com/facts/epl_facebook_table.html [Accessed 22 Mar. 2019].
- [8] The Daily Viz. (2019). *How Common is Your Birthday? This Visualization Might Surprise You*. [online] Available at: <http://thedailyviz.com/2016/09/17/how-common-is-your-birthday-dailyviz/> [Accessed 18 Mar. 2019].