MATHS EXPLORATION

Exploring the Solution to the Birthday Problem

Introduction

My interest in the Birthday problem first came about during an assembly I listened to during my first year at secondary school, a guest speaker came in and took it upon himself to ask us how many people in the audience, we thought, shared a birthday and the probability of that happening, the answers that people came up with were much lower than what the actual answer should've been. People gave outrageous answers like there being no chance of people sharing birthdays, some said it was 1/100, but answers like these were all wrong and that has always stuck in my head, why were these answers wrong at the time? And why were they so far away from the answer the speaker gave us? One of the aspects that makes this problem so intriguing is that the answer is much different than people intuitively expect at first.

In terms of probability theory, the birthday problem or birthday paradox involves the probability that, in a set of *n* randomly chosen people (*n* representing the number of people in the sample), some pair of them will share a birthday together. What I plan on focusing on, is exploring a few of the methods of how people have aimed to solve the birthday paradox, how the variants of the birthday paradox have been able to improve the accuracy of the approximation and also discuss the applications of these to the study of probability more generally.

The probability that at least two people in a room of *n* people share a birthday is 1 minus the probability that there is no match, which can be used to generate the following general formula.

The Theorem

$$\text{P (no birthdays shared)} = \frac{{}^{365}P_n}{365^n}$$

No explanation of notation.

$$\text{P (at least 2 people share a birthday)} = 1 - \text{P (no birthdays shared)}$$

Proving the Birthday Problem

If no two people share a birthday, then the 1st student is capable of choosing a birthday in 365 ways (number of days in a year), the second student has 364, the third has 363 ways, and so on. If there are 23 students in a room, the probability that no two sharing a birthday is:

Theorem:

P(none share the same birthday) = $\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \ldots$

The mathematics requires more clarity and explanation.

P(none share the same birthday) = $\frac{^{365}P_n}{365^n}$

P(at least 2 students share the same birthday) = 1 – P(none share the same birthday)

Worked Example:

P(none share the same birthday) = $\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \ldots \frac{343}{365}$

P(none share the same birthday) = $\frac{^{365}P23}{365^{23}}$

P(at least 1 pair of students share the same birthday) = $1 - \frac{^{365}P_{23}}{365^{23}}$

= 0.507

I used n = 23 because I know it gives close to 0.5 as a probability, this means you'd expect a 50% chance that at least 1 pair of students share the same birthday.

How do you know?

Exactly?

Table 1

P(at least one similar pair)

n	2	3	4	5	10	20	23	30	50
P_n	.003	.008	.016	.027	.117	.411	.507	.706	.970

Source?

Strong Birthday Problem approximation

The Strong birthday problem is a variant of the birthday problem which asks for the probability of a much stronger coincidence than its more basic form. This problem starts to ask what the probability is in a group of *n* people, that everyone in the group shares their birthday with another within that group. The problem also asks what is the number of people needed, to make the probability that everyone has a match equal to or greater than ½? It seems that with more than 3064 people, the probability becomes more than 1/2 that all of them share a common birthda

To start, suppose there are n people, what is the probability that there are exactly i pairs who share different birthdays? So using the strong birthday problem approximation, the probability that exactly i pairs (i = numb of pairs) of equal birthdays is shown as.

$$\frac{365!\,n!}{i!(365-n+i)!2^i(n-2i)!}$$

Probability of what?

Worked Example:

In a group of 3064 people, the probability is greater than half. This is by using an iteration approximation, the act of repeating an approximation with the aim of approaching the desired target of 0.5 in this case or result to make the result more accurate. So an accurate iterative approximation of the smallest n required to make...

$$P_n = 0.5$$

$$= P \text{ is } \frac{n_1}{m}$$

$$= \log\left(\frac{ni-1}{m}\right) + \frac{n1}{m} \; for\ P = 0.5$$

Five iterations give us the value of 3064.

Table 2
P(each person has a shared birthday)

n	*Pn*
2000	.0001
2500	.0678
2700	.1887
2800	.2696
3000	.4458
3063	.4999
3064	.5008
3500	.7883
4000	.9334
4400	.9751

"Under certain configurations of m and n, the number of unique individuals has a Poisson limit distribution. Under other configurations, there can be other limiting distributions. By linking N to the number of cells with exactly one b in a multinomial allocation, the various limiting distributions corresponding to various configurations of m, n can be obtained from the results in Kolchin et al. (1978)."

For general random birthday probabilities, the mean is easy to find for the distribution of *N*.

$$E(N) = n\sum_{k=1}^{m} Pk(1-Pk)^{n-1}$$

Poisson Birthday approximation

The Poisson approximation is yet another variant to the standard birthday problem with an increased accuracy, it displays the probability that at least 1 pair of individuals share a birthday and gives accurate results. The distribution of the number of similar pairs of individuals sharing same birthday is *W*.

Theorem:

If $m,n \longrightarrow \infty$ in a way that $\frac{n(n-1)}{m} \longrightarrow 2\lambda$

m = number of calendar days in a year

n = number of people

Then $W \Longrightarrow \text{Poi}(\lambda)$

'f m=365 and n=23, then $\mathbf{2\lambda} = \frac{23(23-1)}{365}$

$= 1.3863$

$\boldsymbol{\lambda} = 0.69315$

Using the Poisson approximation, we get $P(W \geq 1)$

$= 1 - P(W=0)$

$= 1 - e^{-0.69315}$

$= 0.500001.$

This is a good approximation compared to the standard birthday approximation because it's closer to 0.5 with this approximation.

So when $n = 30$ then this formula gives you 0.696 in this approximation, whereas in the standard birthday problem you get an approximation of 0.706, when trying to get to the number of people where the probability is close to 0.7 as possible.

Applications of the birthday problem

The birthday problem is a probability problem first presented by the mathematician Richard von Mises in 1939, nowadays the birthday problem has applications of it which exist in many fields. One is called Class Phenotype Probability. With six characteristics (e.g. blood type, sex, mid-digital hair positive/negative, earlobes, PTC taste receptor, RH positive/negative), you can be capable of determine the probability that a combination of 2 of them exists and also the probability that two people share a particular combination. This can be useful in medicine when considering the chance of finding matches between prospective donors and recipients. The strong birthday problem can also be applied to real life situations such as in eye-witness testimonies of look-alikes; this would be of interest to criminologists in terms of understanding whether or not an eye-witness testimony of someone who committed a crime could just be the perpetrator's look alike and not the actual criminal himself. In criminal investigations, its often seen that a picture of a suspect circulates in the media drawn on the basis of information provided by an eye-witness based off some of the key physical features of a criminal. Times where the wrong person is apprehended are common, because an innocent person happens to look like the person drawn in the picture. "The various configurations of physical features can be regarded as the cells of a multinomial and people regarded as balls." With this statement, if we were to consider using 10 key physical features, each with three different categories (such as tall, medium, short for height), then we have a multinomial with m = 310 cells. Though if n, the relevant population size is large enough, then the number of cells with 2 or more balls would be large too. This would imply that the person in the picture being circulated may have a look-alike. The calculations in the strong birthday problem have application to criminology, in particular, assessing the likelihood of misapprehension in criminal incidents.

Some attempt at reflection.