Exploring the Birthday Paradox

Introduction:

A few weeks ago at a family event, I found out that one of my uncles in a group of around 30 people had the same birthday as my sister. Having been fascinated with probability theory since high school, I began contemplating the odds of this coincidence occurring. Initially, I assumed that, since in a given year there are 365 days to be born on, the probability of two people sharing their birthdays in a group of 30 people would be really low. However, on further research on this topic, I was introduced to the Birthday Paradox/Problem.

The birthday problem refers to the probability that in a set of *n* random people, two of them will share the same birthday. It belongs in the world of the probability theory and is alternatively known as the birthday paradox. The paradox does not arise from the fact that the solution arrived at through this problem is illogical, but the underestimation of this solution by people, hence perceiving it as a paradox of sorts.

While considering the probability of a person sharing his/her birthday with someone, we instinctively assume that we are that someone, a consideration that makes the scenario seem more improbable than it is. Hence, our assumption is that the probability of two people sharing the same birthday is very low since there are 365 days in a year and therefore 365 possibilities.

Another factor known as the **Pigeonhole Principle** come into play here. The principle states *that "given n boxes and m>n objects, at least one box must contain more than one object."* [1]This means that supposing we have 53 children, at least two of them should have their birthday in the same week (since there are 52 weeks in a year and 53>52). Hence, accordingly, the probability of two people having same birthday will be 100% if there are a total of 366 people (not a leap year).

However, in my exploration of this paradox, the numbers are quite different from what we perceive them to be.

[1] http://mathworld.wolfram.com/DirichletsBoxPrinciple.html. 2:38pm, 15/11/13

Case 1:

In this case we explore the minimum number of people required in a room such that there is a greater than or equal chance of two people having the same birthday than not.

We make two assumptions before going ahead with this problem, and the ones tackled further on:

- There are no leap years. Therefore, the number of days in a year is 365.
- There is an equal probability of a person being born on any date. No particular date has a greater chance of being born on.

Now, let the probability of two people being born on the same date be P(a). Alternatively, let the probability of two people *not* being born on the same day be P(b). Since these events are complementary

$$P(a) + P(b) = 1$$

It is easier to calculate P(b), hence I will determine it first.

We know,

$$Probability = \frac{Desired\ outcomes}{Total\ number\ of\ outcomes}$$

Since there are 365 days in a year, the total number of days in which a person can have his birthday is 365. Also, the 1st person could have been born on any day; therefore the probability of him being born on some day is $\frac{365}{365}$

The 2nd person can only be born on any of the remaining 364 days, since the birthdays cannot be shared. Therefore, for him the probability is $\frac{364}{365}$

Similarly, for the 3rd person the probability is $\frac{363}{365}$, so on and so forth. Hence, for the N-th person, the probability of being born on a day other than the previous N-1 days would be $\frac{365-N+1}{365}$

Therefore, if there are *N* people in the room, the probability of no person sharing his/her birthday is

$$P(b) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \ . \ . \ . \ . \ . \frac{365-N+1}{365} = \frac{365!}{365^N(365-N)!}$$

Hence,

$$P(a) = 1 - \frac{365!}{365^N (365 - N)!}$$

Also, we require the minimum number of people such that P(a) ≥ P(b); therefore, P(a) is atleast greater than or equal to 0.5, since P(a)+P(b)=1. Hence

$$1 - \frac{365!}{365^N (365 - N)!} \geq 0.5$$

$$\frac{365!}{365^N (365 - N)!} \leq 0.5$$

Since it is not possible to further solve this inequality without using a mathematical calculator, we can convert it into a Taylor's series to find an algebraic solution. In the Taylor's series expansion of *e*, the value of e^x is given as

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \qquad -\infty < x < \infty$$

When *x* is small, we can approximate the series to

$$e^x = 1 + x$$

We had determined that

$$P(b) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \; . \; . \; . \; . \; . \frac{365-N+1}{365}$$

This can be rewritten as

$$= \left(1 - \frac{0}{365}\right) \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \left(1 - \frac{3}{365}\right) \; . \; . \; . \left(1 - \frac{N-1}{365}\right)$$

Using these values, we can create a Taylor's series by setting $x = \frac{0}{365}$ for the first term, $x = \frac{-1}{365}$ for the second term and so on and so forth. Therefore, we get

$$P(b) = e^{\frac{0}{365}} \times e^{\frac{-1}{365}} \times e^{\frac{-2}{365}} \times e^{\frac{-3}{365}} . \; . \; . \, e^{\frac{-(N-1)}{365}}$$

But, $e^{\frac{0}{365}}$ is 1, therefore

$$= e^{\frac{-1}{365}} \times e^{\frac{-2}{365}} \times e^{\frac{-3}{365}} \ldots e^{\frac{-(N-1)}{365}}$$

$$= e^{\frac{-1}{365}(1+2+3\ldots N-1)}$$

$$= e^{\frac{-1}{365}(\sum_{i=1}^{N-1} i)}$$

$$= e^{\frac{-1}{365}(\frac{N(N-1)}{2})}$$

Now that we have found P (b), we can use the P (a)+P (b)=1 to find P (a)

$$P(a) = 1 - e^{\frac{-1}{365}(\frac{N(N-1)}{2})}$$

We need P(a) $\geq P$(b), therefore

$$1 - e^{\frac{-1}{365}(\frac{N(N-1)}{2})} \geq 0.5$$

$$e^{\frac{-1}{365}(\frac{N(N-1)}{2})} \leq 0.5$$

Logging both sides,

$$\frac{-1}{365}\left(\frac{N(N-1)}{2}\right) \leq ln\, 0.5$$

$$N(N-1) \geq -730\, ln\, 0.5$$

$$N^2 - N \geq -730\, ln\, 0.5$$

$$N^2 - N + 730\, ln\, 0.5 \geq 0$$

Hence we can use the quadratic formula, $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, to find the value of N

$$N = \frac{1 \pm \sqrt{1 - 4(730 \ln 0.5)}}{2}$$

$$N = \frac{1 \pm 44.999}{2}$$

But N is positive,

$$N = \frac{1 + 44.999}{2}$$

And we need the smallest integer,

$$\therefore N \sim 23$$

The least number of people required in a room such that the probability of two of them having the same birthday is greater than or equal to the probability of each individual being born on a separate date. The answer we arrive at is **23.** Therefore, when there are 23 individuals in the room there is a **50%** chance that two individuals have the same birthday. This number is much smaller than the one we would assume – 183 – using the Pigeonhole principle. It also shows that in the group of 40 people in my family function, the probability of my sister sharing her birthday with my uncle was considerably high, something that I thought was an unlikely event to happen.

It is also possible to make a graph for the probability of two people sharing a birthday. In order to do so we take the equation:

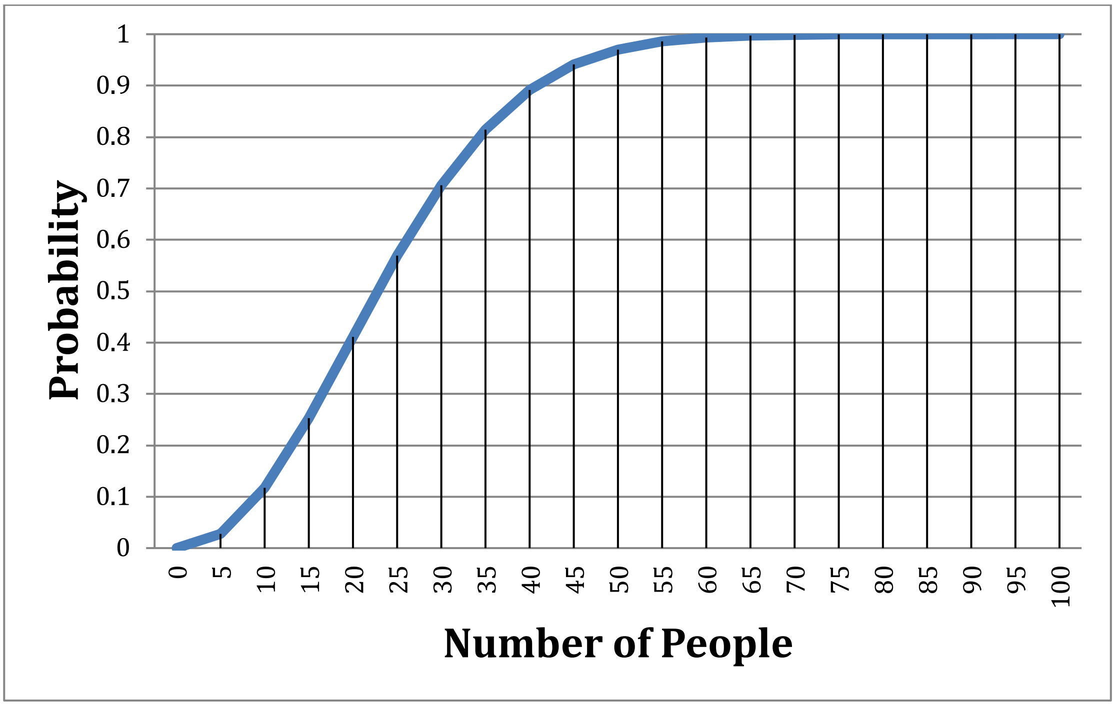
$$P(a) = 1 - e^{\frac{-1}{365}\left(\frac{N(N-1)}{2}\right)}$$

And obtain the following data using a calculator,

Number of People	**Probability (Till 5 decimal spaces)**
0	0
5	0.02714
10	0.11695
15	0.25290
20	0.41144
25	0.56870
30	0.70632

35	0.81438
40	0.89123
45	0.94098
50	0.97037
55	0.98626
60	0.99412
65	0.99768
70	0.99916
75	0.99972
80	0.99991
85	0.99998
90	0.99999
95	1
100	1

Using these values we construct a graph,



In this graph we can see that at approximately 23 individuals the probability of two people sharing their birthday is around 0.5. Interestingly, it can also be seen that in a room of around **57-58** individuals the probability of two people sharing a birthday becomes around **99%** (since a probability of **100%** is the horizontal asymptote, it cannot be achieved), a number that is quite small considering the total number of days in a year. It also shows that at my family function, in that group of 30 people there was 70% chance of people having the same birthdays, which turned out to be the birthdays of my uncle and sister.

There is an alternative method that can be used to find the probability of two people sharing a birthday when the number of people in the room is given. This method can be used to check the answer that is obtained on solving the Taylor series equation for Case 1.

In a room with *r* people, to determine the total number of combinations for pairs of two people, we use the formula

$$C(r,2) = \frac{r!}{2!\,(r-2)!}$$

Therefore for 23 people, the number of combinations is

$$C(23,2) = \frac{23!}{2!\,(21)!}$$

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

$$= 253$$

Hence there are **253** pairs of two people each possible in a room of 23 individuals

In order to calculate P(b), the probability of no person sharing his birthday, let us assume that there are just two people in the room. Therefore, the first person can be born on any of the 365 days, but the second individual must be born on one of the remaining 364 days so that he has a different birthday. Hence, the probability of these two people having different birthdays is

$$1 - \frac{1}{365}$$

$$= \frac{364}{365}$$

Now, for a room of 23 people, in order to determine the probability of a similar event, we will have to multiple the solution we arrived at **C(23,2)** times.

$$P(b) = \frac{364}{365}^{253}$$

$$\sim 0.499523$$

Therefore the probability of two individuals sharing a birthday in a room of 23 people is

$$1 - 0.499523$$

$$\sim 0.50047$$

Hence, the answer we get is approximately **50%,** and it verifies the method we used to solve the problem posed in case 1.

Case 2:

Once I realized that the probability of two people sharing a birthday was higher than 50% at my family function, I began contemplating that what would the size of a group of people have to be such that I have a greater chance of sharing my birthday someone?

Therefore, in this case we explore the minimum number of people required in a room such that there is a greater than or equal chance of a person having the same birthday as me than not.

Before moving on with this problem, lets look at the probability of a person being born on October 12th (my birthday) in a room of 23 people.

Let the probability of none of the 23 people having October 12th as their birthday be Q(b). Since they can be born on all days except one,

$$Q(b) = \left(\frac{364}{365}\right)^{23}$$

Therefore, the probability that at least one person was born on October 12th, Q(a), is

$$Q(a) = 1 - \left(\frac{364}{365}\right)^{23}$$

This is because both events are complementary and hence, Q(a)+Q(b)=1

$$Q(a) = 1 - (0.99726)^{23}$$

$$Q(a) = 1 - 0.93885$$

$$Q(a) = 0.06115$$

Therefore there is roughly a **6%** chance that in a group of 23 individuals at least one of them will share his birthday with me as compared to **50%** for two random individuals to share their birthdays. This explains why we assume that the probability of two people sharing a birthday is really low. We subconsciously believe one of the persons to be us and hence the probability of the event decreases by a great extent, as large as **88%** in this case, leading to the 'paradox' in the birthday problem.

Moving on to the case, we need to determine the minimum number of individuals required such that Q(a) ≥ Q(b). Hence, the 23 people are replaced by *N* number of people.

Therefore, the probability that none of the *N* people were born on October 12th is

$$Q(b) = \left(\frac{364}{365}\right)^N$$

And,

$$Q(a) = 1 - \left(\frac{364}{365}\right)^N$$

Also as in case 1, due to the complementary nature of events, P(a) is at least greater than or equal to 0.5.

$$1 - \left(\frac{364}{365}\right)^N \geq 0.5$$

$$\left(\frac{364}{365}\right)^N \leq 0.5$$

Logging both sides,

$$N \times log\left(\frac{364}{365}\right) \leq log\, 0.5$$

$$-0.002743N \leq -0.693147$$

$$0.002743N \geq 0.693147$$

$$N \geq 252.696$$

$$\therefore N \sim 253$$

Hence, it takes a room of around **253** people so that the probability of at least one person having the same birthday as me is **50%.** It is interesting to note that the reason this number is greater than even half the number of days in a year is because of the chance that other people could have the same birthdays, decreasing the probability of me sharing my birthday with someone.

In fact it is also possible to construct a graph for the probability of a person sharing his birthday with me using the equation,

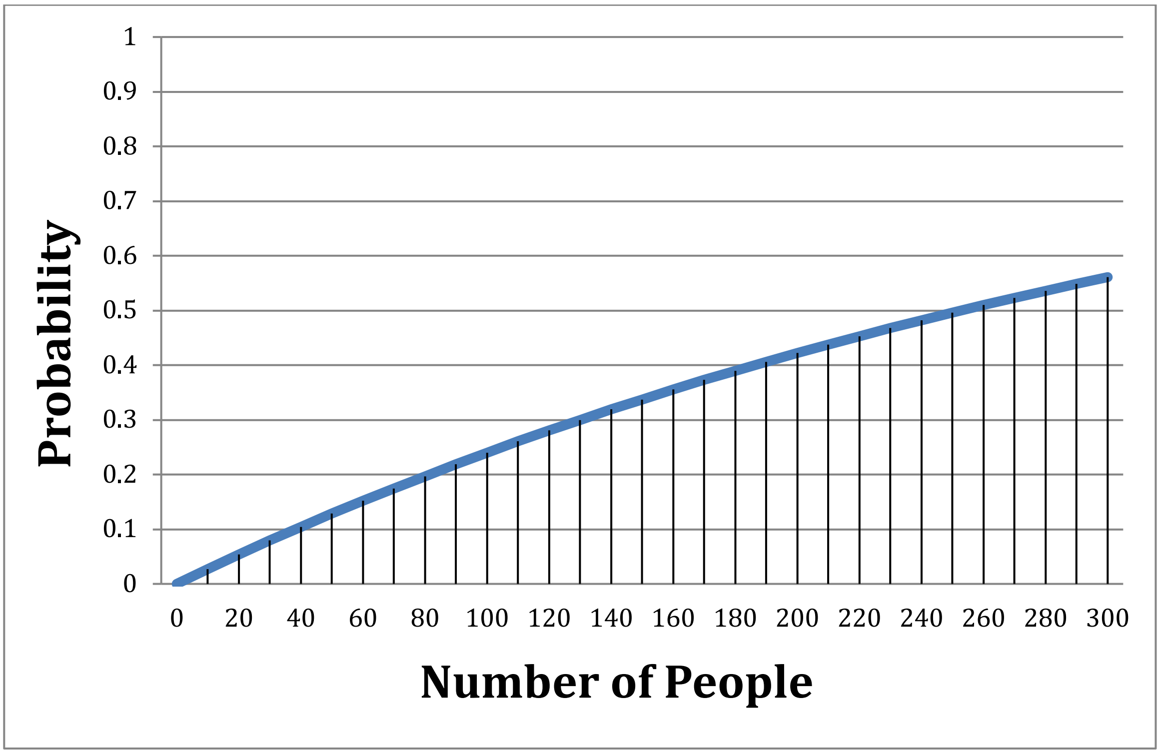
$$Q(a) = 1 - \left(\frac{364}{365}\right)^N$$

And obtain the following data through a calculator,

Number of People	**Probability (Till 5 decimal spaces)**
0	0
10	0.02706
20	0.05339
30	0.07901
40	0.10393
50	0.12818
60	0.15177
70	0.17473
80	0.19706
90	0.21879
100	0.23993
110	0.26050
120	0.28051
130	0.29998
140	0.31893
150	0.33736
160	0.35529

170	0.37274
180	0.38971
190	0.40623
200	0.42230
210	0.43793
220	0.45314
230	0.46794
240	0.48234
250	0.49635
260	0.50998
270	0.52324
280	0.53614
290	0.54869
300	0.56091

Using these values we construct a graph,



This graph makes it apparent as to why it is extremely unlikely to find a person with whom we share our birthday. As calculated earlier, we need a minimum of **253** individuals for a **50%** chance of finding someone with the same birthday as us. In fact, the huge number of people required for a 100% probability, made it impractical for me to even complete the graph. However, if I do continue it, then it will take at least a group of **1680** individuals for me to have a **99%** (as probability of 100% cannot be achieved) chance of finding someone born on October 12th. That is an extremely large number of people, around the size of my entire school!

Extension:

Recently I was standing in a queue to purchase tickets for an upcoming film. I was standing at the fifth position from the front of the line, when the theatre employees announced that the first individual who shares his/her birthday with any of the individuals in the queue in front of him/her would win a free movie ticket. I began contemplating the probability of me being that person, and this problem became my *"movie line dilemma*". In this problem the first individual who has the same birthday with any of the individuals in front of him wins a free movie ticket. Hence, the problem looks at where an individual should stand in line to have the best chance of winning the free movie tickets. It is assumed that the 1st person in line cannot win a movie ticket.

Let *N* be the individual in line purchasing a ticket. Since we the probability of the 1st Individual winning a ticket is zero,

$$P(1) = 0$$

In order for the 2nd individual to win tickets, his birthday would have to be the same as that of the 1st individual. Also the 1st person can have his birthday on any of the 365/365 days, but the 2nd person can only have it on 1/365 days. Therefore,

$$P(2) = \frac{365}{365} \times \frac{1}{365}$$

For the 3rd individual to win, his birthday must be same as that of 1st or 2nd individual. Now the 1st person can be born on 365/365 days, and we assume that the 2nd person did not share his birthday with the 1st person; therefore he can be born on any of the 364/365 days. This leaves the 3rd individual with 2/365 days to be born on in order to win tickets. Hence,

$$P(3) = \frac{365}{365} \times \frac{364}{365} \times \frac{2}{365}$$

Similarly, for the 4th person,

$$P(4) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{3}{365}$$

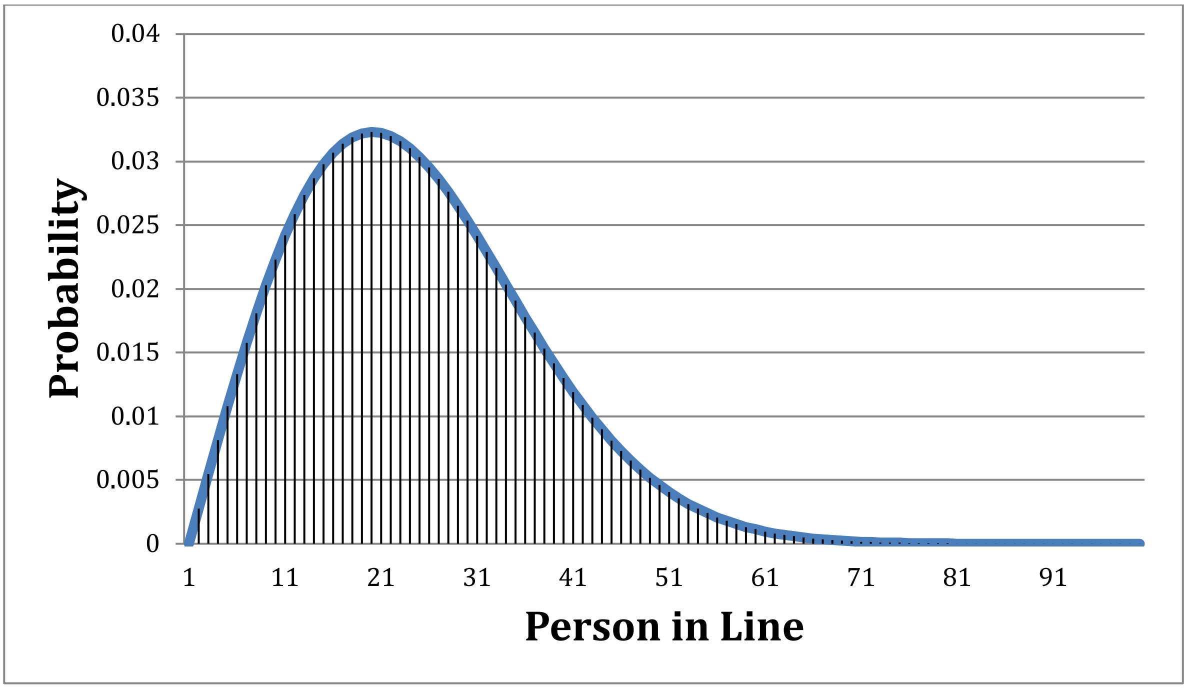
Therefore, we can obtain the general formula,

$$P(N) = \frac{[365 \times 364 \times 363 \ldots . (365 - n + 2)](n-1)}{365^n}$$

Using this equation, we can calculate the following data,

N	Probability	N	Probability	N	Probability	N	Probability
1	0	26	0.02954	51	0.00406	76	0.00006
2	0.00274	27	0.02862	52	0.00357	77	0.00005
3	0.00546	28	0.02760	53	0.00313	78	0.00004
4	0.00815	29	0.02651	54	0.00274	79	0.00003
5	0.01078	30	0.02535	55	0.00239	80	0.00002
6	0.01333	31	0.02414	56	0.00207	81	0.00002
7	0.01577	32	0.02289	57	0.00179	82	0.00001
8	0.01810	33	0.02162	58	0.00154	83	0.00001
9	0.02029	34	0.02035	59	0.00132	84	0.00001
10	0.02232	35	0.01907	60	0.00113	85	0.00001
11	0.02419	36	0.01780	61	0.00097	86	0.00001
12	0.02588	37	0.01655	62	0.00082	87	0.00000
13	0.02739	38	0.01533	63	0.00069	88	0.00000
14	0.02869	39	0.01415	64	0.00059	89	0.00000
15	0.02980	40	0.01301	65	0.00049	90	0.00000
16	0.03070	41	0.01192	66	0.00041	91	0.00000
17	0.03140	42	0.01088	67	0.00034	92	0.00000
18	0.03190	43	0.00989	68	0.00029	93	0.00000
19	0.03221	44	0.00896	69	0.00024	94	0.00000
20	0.03232	45	0.00809	70	0.00020	95	0.00000
21	0.03225	46	0.00728	71	0.00016	96	0.00000
22	0.03201	47	0.00652	72	0.00013	97	0.00000
23	0.03160	48	0.00582	73	0.00011	98	0.00000
24	0.03105	49	0.00518	74	0.00009	99	0.00000
25	0.03036	50	0.00459	75	0.00007	100	0.00000

Using these values we construct a graph,



From both, the data calculated and the graph constructed, it becomes clear that it is somewhere around the **20th** person that the probability of sharing the birthday with a person in front of the line becomes greatest at approximately **3.23%**. This means that if a person wants to have the greatest chance of winning a movie ticket he should stand at the 20th spot from the front of the line. Sadly, for me that day, the probability to win a movie ticket was a dismal **1.078%**.

Conclusion:

This exploration provided me the opportunity to explore the intriguing Birthday Problem. Drawing inspiration from the coincidence of my sister sharing her birthday with my uncle, I used this project to explore the minimum number of people required in a room such that the probability of two people sharing a birthday is 0.5. Investigating the probability of me and another person having the same birthday helped me to truly understand the 'paradox' of the birthday problem. I learnt that the paradox is an amalgamation of psychology and mathematics, stemming from the fact that we subconsciously assume ourselves to be one of the two people sharing birthdays, and thereby distorting the mathematics of the problem. Using the birthday problem to tackle my movie dilemma, gave me a better insight as to where I should stand the next time my local theatre decided to give free tickets again!

In all, this exploration gave me a valuable insight into the world of probability and made me rethink my existing prejudices on the likelihood of a number of events occurring.

Bibliography

- http://mathworld.wolfram.com/DirichletsBoxPrinciple.html. 2:38pm, 15/11/13

- Aldag, Stacey. "A Monte Carlo Simulation of the Birthday Problem." 1 July 2007. Digitalcommons.unl.edu. University of Nebraska-Lincoln. 15 December 2013 <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1027&context=mathmidexppap>.

- Sun, Jared. "The Birthday Paradox: A Simple and Surprising Probability Problem." 3 August 2011. http://cosmos.ucdavis.edu/. University of California, Davis. 15 November 2013 <http://cosmos.ucdavis.edu/archives/2011/cluster6/Sun_Jared.pdf>.

- Erika Batista, Gaël Canal, Karim Ziadeh. "The Birthday paradox: Operational Research and Optimization (Master EEJSI)." December 2012. http://www.eejsi.org. PRES Sorbonne University. University of Paris II. 15 November 2013 <http://www.eejsi.org/M2/claroline195/claroline/backends/download.php?url=L1RlY2huaWNhbF9wcm9qZWN0cy9UaGVfQmlydGhkYXlfcGFyYWRveC5wZGY%3D&cidReset=true&cidReq=THIN1>.