

## Birthday Paradox

Yue Kwok Choy

### (1) Introduction

Please make a guess: *In a class of 23 students, what is the probability that at least two students sharing a birthday, assuming that there are 365 days in a year (forget about the leap year) ?*

This is the classical mathematical problem called the birthday paradox. In fact, this is not really a “paradox”, as paradoxes usually lead to “logical contradictions”. We call this a “paradox” because one not taking proper probability training tends to under-estimating the value of this probability. He may have an experience that it is difficult for *him* to find another one in a small group (such as 23 here) to have the same birthday as his. However, we are not demanding the coincidence between *specific* students and *specific* birthdays, but just *any* two students and *any* birthday in a year. The resulting probability is therefore higher.



The calculation is not difficult. If no two share a birthday, then the first student can choose a birthday in 365 ways, the second student in 364 ways, the third in 363 ways, and so on. If there are 23 students in the room, the probability that no two sharing a birthday is:

$$P(\text{none sharing the same birthday}) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{343}{365}$$

∴ P(at least two students sharing the same birthday)

$$\begin{aligned} &= 1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{343}{365} \\ &= 1 - \frac{365 \times 364 \times 363 \times \dots \times 343}{365^{23}} \approx \underline{\underline{0.507297234}} \quad \dots(1) \end{aligned}$$

There is just more than half the probability that at least two students having the same birthday in a class of just 23 students.

### (2) Factorial and Permutation

If we have  $n$  students in a class instead of 23, then

P(at least two students sharing the same birthday)

$$= 1 - \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{365 - n + 1}{365} \quad \dots(2)$$

If we define, for any natural number  $n$ , the *factorial* of  $n$  :

$$n! = n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1 \quad \text{and} \quad 0! = 1$$

then we all know that  $n!$  is the number of ways to arrange  $n$  different objects in a sequence.

Equation (2) can be rewritten in the form:

$$\begin{aligned}
 &P(\text{at least two students sharing the same birthday}) \\
 &= 1 - \frac{365!}{365^n(365 - n)!} \qquad \dots(3)
 \end{aligned}$$

If we further define, for non-negative integers  $n$  and  $r$ ,  $n \geq r \geq 0$ ,

$$\text{the permutation: } P_r^n = \frac{n!}{(n-r)!}$$

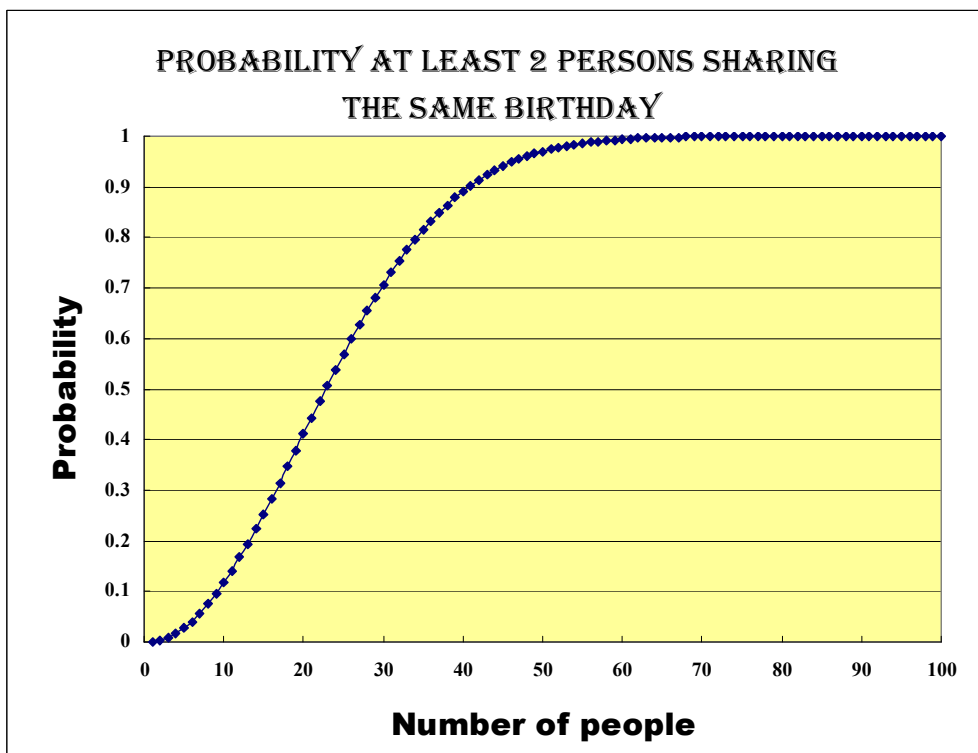
then we know that  $P_r^n$  gives us the number of ways in choosing  $r$  objects out of  $n$  different objects and arranging these  $r$  objects in a sequence in which order is **important**.

Equation (3) can be further rewritten in the form:

$$\begin{aligned}
 &P(\text{at least two students sharing the same birthday}) \\
 &= 1 - \frac{P_n^{365}}{365^n} \qquad \dots(4)
 \end{aligned}$$

### (3) Combination

How can we know that 23 students are enough to get a probability of just over 0.5 with at least 2 students sharing the same birthday? One method is to use a computer program and find the corresponding probability for  $n = 2, 3, 4, \dots$  onwards until you can get a value of probability more than 0.5. A small EXCEL spreadsheet is attached for this purpose.



Another method which is more mathematical is to solve for a smallest integer  $n$ ,

$$1 - \frac{P_n^{365}}{365^n} > 0.5 \quad \text{or} \quad \frac{P_n^{365}}{365^n} < 0.5 \quad \text{or} \quad P_n^{365} < 0.5(365^n)$$

However, this is difficult to solve.

So if we define, for non-negative integers  $n$  and  $r$ ,  $n \geq r \geq 0$ ,

$$\text{the combination : } C_r^n = \frac{n!}{r!(n-r)!}$$

then we know that  $C_r^n$  gives us the number of ways in choosing  $r$  objects out of  $n$  different objects and arranging these  $r$  objects in a sequence in which order is **unimportant**.

The probability of any two students not having the same birthday is  $\frac{364}{365}$ . In a class

containing  $n$  students, there are  $C_2^n = \frac{n(n-1)}{2}$  pairs of students. The probability of no two students sharing the same birthday can be approximated by assuming that these events are *independent* and hence by multiplying their probability together. Therefore we have

$$P(\text{at least two students sharing the same birthday}) \approx 1 - \left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}} \quad \dots(5)$$

It is really an approximation as the pairs  $(A, B)$  and  $(B, C)$  having common birthdays does not imply that  $(A, C)$  have no common birthday. However, the probability in (5) is a good approximation if  $n$  is getting bigger and the number of pairs is even bigger.

So if we want to find the smallest positive integer  $n$  to have a probability of 0.5 we solve:

$$1 - \left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}} > 0.5 \quad \text{or} \quad \left(\frac{364}{365}\right)^{\frac{n(n-1)}{2}} < 0.5 \quad \text{or} \quad \left(\frac{365}{364}\right)^{\frac{n(n-1)}{2}} > 2$$

Taking logarithm, we get :

$$\frac{n(n-1)}{2} \log\left(\frac{365}{364}\right) > \log 2 \quad \text{or} \quad \frac{n(n-1)}{2} < \log\left(\frac{365}{364}\right) / \log 2 \approx 505.3039777$$

Hence, we need to solve the quadratic inequality:

$$n^2 - n - 505.3039777 < 0 \quad \dots(6)$$

Solving (6), using quadratic equation formula to half, we get:

$$n < -21.98452752 \quad \text{or} \quad n > 22.98452752 \quad \dots(7)$$

Since  $n$  is the *smallest positive integer* satisfying (7), we get  $n = 23$ .

#### (4) Three students sharing birthday

It is rather difficult to calculate the probability that at least 3 students sharing a birthday in a class of  $n$  students. We find out the probability that no 3 students sharing a birthday first and the remainder is what we want.



(a)  $\Omega = \{\text{all possible combination for } n \text{ students}\}$

$$N(\Omega) = 365^n$$

(b)  $A = \{\text{none of } n \text{ students having the same birthday}\}$

$$N(A) = P_n^{365}$$

(c) (i)  $B_1 = \{1 \text{ pair of students (2 students) having the same birthday}\}$

$$N(B_1) = C_2^n P_{n-1}^{365}$$

(ii)  $B_2 = \{2 \text{ pairs of students (i.e. 4 students) having the same birthday}\}$

$$N(B_2) = \frac{C_2^n C_2^{n-2} P_{n-2}^{365}}{2!}$$

(iii)  $B_3 = \{3 \text{ pairs of students (i.e. 6 students) having the same birthday}\}$

$$N(B_3) = \frac{C_2^n C_2^{n-2} C_2^{n-4} P_{n-3}^{365}}{3!}$$

(iv) If we continue in this way we get the last possible number of pairs:

$B_{\lfloor n/2 \rfloor} = \{\lfloor n/2 \rfloor \text{ pairs of students (i.e. } 2 \times \lfloor n/2 \rfloor \text{ students) having the same birthday}\}$   
where  $\lfloor n/2 \rfloor$  stands for the largest possible integer smaller than  $n/2$ .

$$N(B_{\lfloor n/2 \rfloor}) = \frac{C_2^n C_2^{n-2} C_2^{n-4} \dots C_2^{n-2\lfloor n/2 \rfloor-1} P_{n-\lfloor n/2 \rfloor}^{365}}{\lfloor n/2 \rfloor!}$$

(v)  $B = \{\text{only pairs students having the same birthday}\}$

By summing up all the possible pairs listed in (c) (i) to (iv), we get:

$$N(B) = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{C_2^n C_2^{n-2} C_2^{n-4} \dots C_2^{n-2(i-1)} P_{n-i}^{365}}{i!} = \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{n! 365!}{2^i i! (365-n+i)!} \quad \dots(8)$$

(d)  $C = \{\text{at least 3 students having the same birthday}\}$

$$P(C) = 1 - P(A) - P(B) = 1 - \frac{P_n^{365}}{365^n} - \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{C_2^n C_2^{n-2} C_2^{n-4} \dots C_2^{n-2(i-1)} P_{n-i}^{365}}{i!}$$

$$= 1 - \frac{P_n^{365}}{365^n} - \sum_{i=1}^{\lfloor n/2 \rfloor} \frac{n! 365!}{2^i i! (365-n+i)!} \quad \dots(9)$$

## (5) Multinomial Coefficient

The multinomial coefficient is a generalization of the binomial coefficient.

Let  $k_1, k_2, \dots, k_m$  be non-negative integers giving a partition of  $n$ :  
 $k_1 + k_2 + \dots + k_m = n$ . The multinomial coefficient is defined by  $\frac{n!}{k_1!k_2!\dots k_m!}$ .

This coefficient gives you the number of permutations of  $n$  objects, in which  $k_1$  are alike,  $k_2$  are alike,  $\dots$ ,  $k_m$  are alike.

The birthday paradox can then be solved with the help of multinomial theorem using marbles-in-jars model (MJ model). (Nothing connected with Michael Jackson here.☺)

First we define :

$x : \boxed{y, \dots, y}$  as there are  $x$  jars and there are  $y$  alike marbles inside each jar.



So suppose we like to find the probability of having *exactly* 2 students with the same birthdays and the rest of the students are of different birthdays in a class of 23 students. The corresponding MJ model is :

$$1 : \boxed{2}, \quad 21 : \boxed{1, \dots, 1}, \quad 343 : \boxed{0, \dots, 0}$$

(Interpretation : 1 jar with 2 alike marbles, 21 jars with 1 marbles, 343 jars with no marble.)

Calculation of the probability involves 3 steps:

- (1) Calculation of the number of ways in arranging the **Jars**, here we have:

$$J = \frac{365!}{1! 2! 343!}$$

- (2) Calculation of the number of ways in arranging the **Marbles** in the jars, here :

$$M = \frac{23!}{2! (1!)^{21} (0!)^{343}}$$

- (3) Find MJ and divide by the total number of ways of arranging 23 marbles in 365 jars, that is  $365^{23}$ , here finally we get :

P(one and only one pair of same birthday)

$$= \left[ \frac{23!}{2! (1!)^{21} (0!)^{343}} \right] \left( \frac{365!}{1! 2! 343!} \right) \left( \frac{1}{365^{23}} \right)$$

$$\approx \underline{\underline{0.36344216}} \quad \dots(10)$$

This probability is of course smaller than the probability listed in (1).

**(6) More calculations with MJ model**

(a) Probability of no students with same birthday in a class of 40:

$$\text{MJ model : } 40 : \boxed{1, \dots, 1}, \quad 325 : \boxed{0, \dots, 0}$$

Remember that Probability = MJ / N( $\Omega$ )

P(no students with same birthday)

$$= \left[ \frac{40!}{(1!)^{40} (0!)^{325}} \right] \left( \frac{365!}{40! 325!} \right) \left( \frac{1}{365^{40}} \right)$$

$$\approx \underline{\underline{0.10876819}}$$

Also, we can calculate:

$\therefore$  P(at least two students sharing the same birthday in a class of 40)

$$\approx 1 - 0.10876819 \approx \underline{\underline{0.89123191}} \quad (\text{Quite high!})$$

(b) Probability of *exactly* two pairs of students sharing same birthday in a class of 40:

$$\text{MJ model : } 2 : \boxed{2, 2}, \quad 36 : \boxed{1, \dots, 1}, \quad 327 : \boxed{0, \dots, 0}$$

P(exactly two pairs of students sharing same birthday)

$$= \left( \frac{40!}{(2!)^2 (1!)^{36} (0!)^{327}} \right) \left( \frac{365!}{2! 36! 327!} \right) \left( \frac{1}{365^{40}} \right)$$

$$\approx \underline{\underline{0.27974123}}$$

(c) Probability of *exactly* three students sharing same birthday in a class of 40:

$$\text{MJ model : } 1 : \boxed{3}, \quad 37 : \boxed{1, \dots, 1}, \quad 327 : \boxed{0, \dots, 0}$$

P(exactly 3 students with same birthday, no pair sharing same birthday)

$$= \left( \frac{40!}{(3!)(1!)^{37} (0!)^{327}} \right) \left( \frac{365!}{1! 37! 327!} \right) \left( \frac{1}{365^{40}} \right)$$

$$\approx \underline{\underline{0.010080765}}$$

- Harder Exercise**
- (1) Use the MJ model to work out (9).
  - (2) Use EXCEL or any software to make a graph for (9) for  $n = 3, 4, \dots, 200$ .  
Send to Mr. Yue if you like.