

The Birthday Paradox

jacques@ucsd.edu

Remarks. *These notes should be considered as part of the lectures. For proper treatment of the birthday paradox, the details are written here in full. These notes should be read in conjunction with Lectures 5 and 6, and after the [multiplication principle](#).*

1. The bet. In class I bet that out of the 42 people in attendance, two would have the same birth day and birth month. This may seem a bit of a foolish bet, since there are 365 possibly days in the year, and only 42 people on the class. If there were more than 365 people in the class, then a win would be assured simply because there must be two people with the same birthday. However, with a bit of combinatorics, we can see that in fact there is a more than 90 percent chance that I bet correctly. This is a counter-intuitive fact which is known more generally as the [birthday paradox](#). I put the results of the (correct as it turns out) bet on the course website.

2. Birthdays on any planet. Let us generalize the problem a little bit and write it in terms of probability measure. Suppose we are on a planet with N days in the year, and we want to know the probability that in a sequence of n uniformly chosen people on that planet, at least two have the same birthday. So the sample space Ω is the set of all sequences of n birthdays (there are N^n such sequences so $|\Omega| = N^n$). The probability measure here is the uniform measure on Ω :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{N^n}.$$

for every event $A \subseteq \Omega$. The event we want is the set A of sequences of n different birthdays, because then $P(\bar{A}) = 1 - P(A)$ is the probability that at least two birthdays are the same, by the complement rule.

3. The exact probability. By the multiplication principle,

$$|A| = N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot (N - n + 1).$$

A short way of writing this is using [product \$\Pi\$ notation](#):

$$|A| = \prod_{i=1}^n (N - i + 1).$$

This means multiply out $N - i + 1$ for values of i from 1 to n , and it is very similar to the [sum \$\Sigma\$ notation](#) which you will have seen before in calculus and earlier in the course. Then

$$P(A) = \frac{|A|}{N^n} = \frac{1}{N^n} \prod_{i=1}^n (N - i + 1) = \prod_{i=1}^n \left(1 - \frac{i-1}{N}\right).$$

This is the probability that the bet fails (all birthdays different). Note that we divided each of the n terms in the first product by N to get the second product. We want to let $n = 42$ and $N = 365$. However, the product is hard to compute for such large n , so instead we find an upper estimate for the product: that is we show that the product is in fact less than 0.1.

4. A fact from calculus. To find an upper estimate for $P(A)$, we use the following fact from calculus:

Fact 1. *If x_1, x_2, \dots, x_n are any real numbers, then*

$$\prod_{i=1}^n (1 - x_i) \leq e^{-\sum_{i=1}^n x_i}.$$

Proof of Fact 1. Well the right hand side of the inequality is a product

$$e^{-x_1} \cdot e^{-x_2} \dots e^{-x_n}.$$

The line $y = 1 - x$ is tangent to $y = e^{-x}$ at $x = 0$, and otherwise lies below the curve $y = e^{-x}$, and therefore $1 - x \leq e^{-x}$ for any real number x . Applying this to each e^{-x_i} , we get

$$e^{-x_1} \cdot e^{-x_2} \dots e^{-x_n} \geq (1 - x_1)(1 - x_2) \dots (1 - x_n) = \prod_{i=1}^n (1 - x_i)$$

and this proves Fact 1.

5. Sum of first n integers. The next fact we need is the well-known formula for the sum of the first n integers:

Fact 2. *For a natural number n ,*

$$\sum_{i=1}^n i = 1 + 2 + \dots + n = \frac{1}{2}n(n + 1).$$

Proof of Fact 1. Gauss had a beautiful way of doing this sum: it is half the area of a rectangle with side lengths n and $n + 1$. To see this, multiply the sum by two. Then we can imagine computing the sum in pairs: $1 + n$ and then $2 + (n - 1)$ all the way up to $n + 1$. But the pairs all add up to $n + 1$, and there are n of them, so twice the given sum is $n(n + 1)$, the area of the rectangle shown below. Therefore the sum itself, which is the total area of the grey blocks in the picture, is $n(n + 1)/2$, as required.

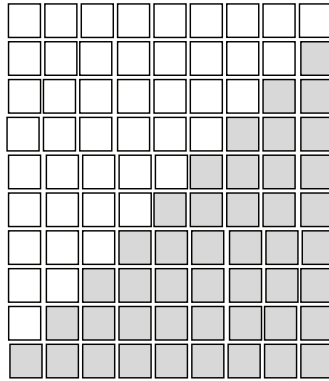


Figure : Gauss' counting

6. The final upper estimate. Using Facts 1 and 2, we find an upper estimate for $P(A)$ as follows (we will call it Fact 3). Let $x_i = (i - 1)/N$ for $i = 1, 2, \dots, n$. Then by Fact 1,

$$P(A) \leq e^{-\sum_{i=1}^n x_i} \leq e^{-\frac{1}{N} \sum_{i=1}^n (i-1)}$$

Using Fact 2 in the exponent,

$$\sum_{i=1}^n (i - 1) = \sum_{i=1}^{n-1} i = \frac{1}{2}(n - 1)n.$$

Therefore our general upper estimate for $P(A)$ is as follows:

Fact 3. *Let N and n be natural numbers. Then*

$$P(A) \leq e^{-\frac{1}{N} \sum_{i=1}^n (i-1)} = e^{-\frac{n(n-1)}{2N}}.$$

7. Back to earth. Consider finally the case of $n = 42$ people and $N = 365$ possible birthdays. By Fact 3,

$$P(A) \leq e^{-\frac{1722}{730}} \leq 0.095 \dots < 0.1.$$

This confirms that there is more than a 90 percent chance that at least 2 out of 42 people have the same birthday, as

$$P(\bar{A}) = 1 - P(A) > 0.9$$

so the original bet had more than a 90 percent chance of being correct.

8. Concluding Remarks. The estimate in Fact 3 says that in general, if n is much more than \sqrt{N} then it is very likely that at least two people will have the same birthday, since $n(n-1)/N$ becomes large when n is much larger than \sqrt{N} . In the course, we will be making this kind of asymptotic statement more precise by giving [limit theorems](#). Exercises related to the birthday paradox can be found in Exercise Sheet 2. You may be asked in homework or exams to write down the exact probability for a birthday-type problem (Step 3 above), but you would not be asked to derive (Steps 4–6) the upper estimate. You may be given the upper estimate and then asked to estimate a probability (as in Step 7).