

Solution

Week 46 (7/28/03)

The birthday problem

- (a) Given n people, the probability, P_n , that there is *not* a common birthday among them is

$$P_n = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right). \quad (1)$$

The first factor is the probability that two given people do not have the same birthday. The second factor is the probability that a third person does not have a birthday in common with either of the first two. This continues until the last factor is the probability that the n th person does not have a birthday in common with any of the other $n - 1$ people.

We want $P_n < 1/2$. If we simply multiply out the above product with successive values of n , we find that $P_{22} = .524$, and $P_{23} = .493$. Therefore, there must be at least 23 people in a room in order for the odds to favor at least two of them having the same birthday.

REMARK: This answer of $n = 23$ is much smaller than most people expect, so it provides a nice betting opportunity. For $n = 30$, the odds of a common birthday increase to 70.6%, and most people still find it hard to believe that among 30 people there are probably two who have the same birthday. The table below lists various values of n and the probabilities, $1 - P_n$, that at least two people have a common birthday.

n	10	20	23	30	50	60	70	100
$1 - P_n$	11.7%	41.1%	50.7%	70.6%	97.0%	99.4%	99.92%	99.9994%

Even for $n = 50$, most people would be happy to bet, at even odds, that no two people have the same birthday. If they seem a bit hesitant, however, you can simply offer them the irrefusable odds of 10 to 1. (That is, you put down, for example, \$10 on the table, and they put down \$1. If there is a common birthday, you take all the money. If there is no common birthday, they take all the money.) Since there is a 97.0% chance of a common birthday among 50 people, a quick calculation shows that you will gain, on average, 67 cents for every \$10 you put down.

One reason why many people do not believe the $n = 23$ answer is correct is that they are asking themselves a different question, namely, "How many people need to be present for there to be a 1/2 chance that someone else has *my* birthday?" The answer to this question is indeed much larger than 23. The probability that no one out of n people has a birthday on a given day is $(1 - 1/365)^n$. For $n = 252$, this is just over 1/2. And for $n = 253$, it is just under 1/2. Therefore, you need to come across 253 other people in order to expect that at least one of them has *your* birthday.

- (b) **First Solution:** Given n people, and given N days in a year, the reasoning in part (a) shows that the probability that no two people have the same birthday is

$$P_n = \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{n-1}{N}\right). \quad (2)$$

If we take the natural log of this equation and use the expansion,

$$\ln(1 - x) = -(x + x^2/2 + \dots), \quad (3)$$

then the requirement $P_n \leq 1/2$ becomes

$$-\left(\frac{1}{N} + \frac{2}{N} + \dots + \frac{n-1}{N}\right) - \frac{1}{2}\left(\frac{1}{N^2} + \frac{4}{N^2} + \dots + \frac{(n-1)^2}{N^2}\right) - \dots \leq -\ln 2. \quad (4)$$

Using the sums,

$$\sum_1^n k = \frac{n(n+1)}{2}, \quad \text{and} \quad \sum_1^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad (5)$$

we can rewrite eq. (4) as

$$\frac{n(n-1)}{2N} + \frac{n(n-1)(2n-1)}{12N^2} + \dots \geq \ln 2. \quad (6)$$

For large N , the first term will be of order 1 when $n \approx \sqrt{N}$, in which case the second- and higher-order terms are negligible. Therefore, keeping only the first term (which behaves like $n^2/2N$), we find that P_n is equal to $1/2$ when

$$n \approx \sqrt{2 \ln 2} \sqrt{N}. \quad (7)$$

Let's look at a few cases:

- For $N = 365$, eq. (7) gives $n = 22.5$. Since we must have an integral number of people, this agrees with the exact result, $n = 23$.
- For $N = 24 \cdot 365 = 8760$ (that is, for births in the same hour), we find $n = 110.2$. This agrees with the exact result, $n = 111$, obtained by multiplying out eq. (2).
- For $N = 60 \cdot 24 \cdot 365 = 525,600$ (that is, for births in the same minute), we find $n = 853.6$. This agrees with the exact result, $n = 854$, obtained by multiplying out eq. (2) (not by hand!). This is a very small number compared to the more than half a million minutes in a year.

REMARKS:

- (1) If we wish to ask how many people need to be in a room in order for the probability to be at least p that some two have the same birthday, then the above derivation is easily modified to yield

$$n \approx \sqrt{2 \ln\left(\frac{1}{1-p}\right)} \sqrt{N}. \quad (8)$$

- (2) The alternative question introduced in part (a), "How many people need to be present for there to be a $1/2$ chance that someone else has *my* birthday?", may also be answered in the large- N limit. The probability that no one out of n people has a birthday on a given day is

$$\left(1 - \frac{1}{N}\right)^n = \left(1 - \frac{1}{N}\right)^{N(n/N)} \approx e^{-n/N}. \quad (9)$$

Therefore, if $n > N \ln 2$, you can expect that at least one of the n people has your birthday. For $N = 365$, we find that $N \ln 2$ is slightly less than 253, so this agrees with the result obtained in part (a). Note that this result is linear in N , whereas the result of the original problem in eq. (7) behaves like \sqrt{N} . The reason for this square-root behavior can be seen in the next solution.

Second Solution: Given a large number, n , of people, there are $\binom{n}{2} = n(n-1)/2 \approx n^2/2$ pairs of people. The probability that a given pair has the same birthday is $1/N$, so the probability that they do not have the same birthday is $1 - 1/N$.¹ Therefore, the probability that no pair has a common birthday is

$$P_n = \left(1 - \frac{1}{N}\right)^{n^2/2} \approx \left(1 - \frac{1}{N}\right)^{N(n^2/2N)} \approx e^{-n^2/(2N)}. \quad (10)$$

This equals 1/2 when

$$n \approx \sqrt{2 \ln 2} \sqrt{N}. \quad (11)$$

in agreement with eq. (7).

REMARK: We assumed above that all the pairs are independent, as far as writing down the $1 - 1/N$ probability goes. Let us now show that this is approximately true. We will show that for large N and n , the assumptions on the coincidence of birthdays in some pairs do not significantly affect the probability of coincidence in other pairs. More precisely, the relation $P_n \approx e^{-n^2/(2N)}$ is true if $n \ll N^{2/3}$. The reasoning is as follows.

Consider n people and N days in a year. Assumptions on the coincidence of birthdays in some pairs will slightly affect the probability of coincidence in other pairs, because the given assumptions restrict the possible birthdays of these other pairs. For example, if it is given that A and B do not have the same birthday, and also that B and C do not, then the probability that A and C do not have the same birthday is $1 - 1/364$ (instead of $1 - 1/365$), because they are both restricted from having a birthday on B 's birthday, whatever it may be.

The maximal restriction that can be placed on the possible birthdays of a given pair occurs when it is known that neither of them has a birthday in common with any of the other $n-2$ people. In this case, the possible birthdays for the last two people are restricted to $N - (n-2)$ days of the year. Therefore, the probability that the last pair does not have a common birthday is $1 - 1/(N - n + 2) > 1 - 1/(N - n)$. This is the most that any such probability can differ from the naive $1 - 1/N$. Therefore, we may say that

$$\begin{aligned} \left(1 - \frac{1}{N-n}\right)^{n^2/2} &\leq P_n \leq \left(1 - \frac{1}{N}\right)^{n^2/2} \\ \implies e^{-n^2/2(N-n)} &\leq P_n \leq e^{-n^2/(2N)}. \end{aligned} \quad (12)$$

The ratio of these upper and lower bounds on P_n is

$$\exp\left(-\frac{n^2}{2N} + \frac{n^2}{2(N-n)}\right) = \exp\left(\frac{n^3}{2N(N-n)}\right). \quad (13)$$

¹This is not quite correct for all the pairs, because two pairs are not independent if, for example, they share a common person. But it is accurate enough for our purposes in the large- N limit. See the remark below.

For large N , this ratio is essentially equal to 1, provided that $n \ll N^{2/3}$. Therefore, $P_n \approx e^{-n^2/(2N)}$ if $n \ll N^{2/3}$. And since the result in eq. (11) is in this realm, it is therefore valid.

Extension: We can also ask the following question: How many people must be in a room in order for the probability to be greater than 1/2 that at least b of them have the same birthday? (Assume that there is a very large number, N , of days in a year, and ignore effects that are of subleading order in N .)

We can solve this problem in the manner of the second solution above. Given a large number, n , of people, there are $\binom{n}{b}$ groups of b people. This is approximately equal to $n^b/b!$ (assuming that $b \ll n$).

The probability that a given group of b people all have the same birthday is $1/N^{b-1}$, so the probability that they do not all have the same birthday is $1 - (1/N^{b-1})$.² Therefore, the probability, $P_n^{(b)}$, that no group of b people all have the same birthday is

$$P_n^{(b)} \approx \left(1 - \frac{1}{N^{b-1}}\right)^{n^b/b!} \approx e^{-n^b/b!N^{b-1}}. \quad (14)$$

This equals 1/2 when

$$n \approx (b! \ln 2)^{1/b} N^{1-1/b}. \quad (15)$$

REMARKS:

- (1) Eq. (15) holds in the large- N limit. If we wish to make another approximation, that of large b , we see that the quantity $(b! \ln 2)^{1/b}$ goes like b/e , for large b . (This follows from Sterling's formula, $m! \approx m^m e^{-m} \sqrt{2\pi m}$.) Therefore, for large N , n , and b (with $b \ll n \ll N$), we have $P_n^{(b)} = 1/2$ when

$$n \approx (b/e) N^{1-1/b}. \quad (16)$$

- (2) The right-hand side of equation eq. (15) scales with N according to $N^{1-1/b}$. This means that if we look at the numbers of people needed to have a greater than 1/2 chance that pairs, triplets, etc., have common birthdays, we see that these numbers scale like

$$N^{1/2}, N^{2/3}, N^{3/4}, \dots \quad (17)$$

For large N , these numbers are multiplicatively far apart. Therefore, there are values of n for which we can say, for example, that we are virtually certain that there are pairs and triplets with common birthdays, but also that we are virtually certain that there are no quadruplets with a common birthday. For example, if $n = N^{17/24}$ (which satisfies $N^{2/3} < n < N^{3/4}$), then eq. (14) shows that the probability that there is a common birthday triplet is $1 - e^{-\frac{1}{6}N^{1/8}} \approx 1$, whereas the probability that there is a common birthday quadruplet is $1 - e^{-\frac{1}{24}N^{-1/6}} \approx \frac{1}{24}N^{-1/6} \approx 0$.

²Again, this is not quite correct, because the groups are not all independent. But I believe it is accurate enough for our purposes in the large- N limit. I have a proof which I think is correct, but which is very ugly. If anyone has a nice clean proof, let me know.