2-2001

A Bonanza of Birthday Bewilderments

Dale K. Hathaway
Olivet Nazarene University, hathaway@olivet.edu

DALE K. HATHAWAY

A Bonanza of Birthday Bewilderments

We are fascinated by coincidences. A pair of airplane hijackings to Cuba take place within 24 hours of each other, the same machine in a plant breaks down twice within two days while other machines continue to work smoothly, or in your advanced Calculus class of 25 there are two students with the same birthday. The culprit here is the birthday problem, often stated as follows: What is the smallest number of people needed to ensure that the probability of at least two having the same birthday exceeds 50%? Typically three assumptions are made: the birthdays are assumed to be equally likely, independent of one another, and the possibility of a February 29 birthday is usually ignored.

The traditional approach to the birthday problem is to consider the complement of the probability of at least one match, i.e., the probability of no matches. There's a good reason for tackling the problem in this apparently backwards way: there are a great number of different types of matches that can happen, there could be a single pair, or two pairs, there could be a triple or four of a kind and a pair. On the other hand, it is straightforward to count the number of ways for there to be no matches. For the first person there are 365 days that could be his birthday, for the second person there are only 364 available days since one day has already been used by the first person, for the third person there are only 363 available days since the first two individuals have already used two days, and so on until the nth person who has $366 - n$ available days. The number of ways to choose n distinct dates from the 365 days of the year is $365 \cdot 364 \cdot 363 \cdots (366 - n)$. The total number of possible sets of birthdays if we allow duplication is 365^n, since each of the n people has 365 available days on which to be born. Together these give the following probability:

$$P(\text{no matching birthdays}) = \frac{365 \cdot 364 \cdot 363 \cdots (366-n)}{365^n}.$$

The complement of this probability is the one we are interested in

$$P(\text{at least one match}) = 1 - \frac{365 \cdot 364 \cdot 363 \cdots (366-n)}{365^n}.$$

The smallest n which gives a probability greater than 0.5 is $n = 23$, which can be checked by simply doing the computation.

Here's another way to think about the problem: for each pair of individuals we have either a match or no matches. There are $n(n-1)/2$ pairs of individuals. For each pair of individuals the birthdays either match, $P(\text{a match}) = 1/365$, or they do not match, $P(\text{no match}) = 364/365$. Thus, the probability of no matches is $(1/365)^0 \cdot (364/365)^{n(n-1)/2}$. The smallest value of n that makes this probability less than .5 is 23. This is beautifully simple but, unfortunately, wrong. The difficulty is that these pairs are not *independent*, i.e., if we know that A and B have the same birthday and that B and C have the same birthday, then we know how A and C are related. (Technically speaking, we're using the binomial distribution which requires independent trials.) Is it amazing that we got the right answer using the wrong method? Not really, it turns out that the binomial distribution in this case is a reasonably good approximation to the true distribution. (Although one could just pretend that this is a truly astounding meta-coincidence, i.e., a coincidence encountered in the study of coincidences.)

The Birthmate Problem

We are even more fascinated with coincidences when they involve us personally. You might wonder how many people it would take to give a better than 50% chance of a match of your own birthday. This problem is called the birthmate problem and is commonly stated: what is the smallest number of people needed so that the probability of a match of a specific predetermined birthday is greater than 50%?

DALE K. HATHAWAY is Professor of Mathematics at Olivet Nazarene University in Illinois.

You might initially think that the answer to the birthmate problem is around 183. That answer would be correct if the birthdays were chosen without replacement, but asking different people their birthdays is actually choosing the birthdays with replacement. We already know that among 23 people there is a better than 50% chance of at least one match, so it would make sense that more than 183 people are needed because there will probably be several duplicated birthdays before we find a match to the specific birthday we are interested in.

To solve this problem consider that each person has a $1/365$ chance of matching the specific birthday. The probability of no matches among n individuals is $(364/365)^n$ since the birthdays are independent. The probability of at least one match of the specific birthday among n individuals is $1-(364/365)^n$. The smallest n for which this probability is greater than 0.5 is $n = 253$. Interestingly enough if we were to allow February 29 birthdays to exist, then the critical value of n to match non-February 29 birthday is still $n = 253$. The probability of a single person having a birthday match of a non-February 29 birthday is $4/(4 \cdot 365 + 1) = 4/1461$, since there is one February 29 birthday every 4 years while all of the other birthdays occur four times. Using the same approach as above we obtain a probability of at least one match out of n people as $1-(1457/1461)^n$. The smallest n that gives a better than 50% probability is $n = 253$. Can you determine the number of people needed to give a better than 50% chance at matching a birthday of February 29?

Birthday Opportunities

While the birthday problem and the birthmate problem are definitely different they do seem to be related. Just how related they are can be examined by considering how many opportunities for a match there are in each case.

For the birthmate problem, the number of opportunities is simply the number of individuals involved. Each individual has one chance to match the specific birthday. So for $n = 253$ there are 253 opportunities for a match.

For the birthday problem each individual has a chance to match all of the birthdays for the individuals already included. The first person provides one possible birthday to match, the second person provides one opportunity for a match with the first person, the third person provides two opportunities for a match, one with each of the first two individuals, and so on. After the nth person has been added there are $1 + 2 + 3 + \cdots + n - 1 = n(n-1)/2$ opportunities for a match. Amazingly, for $n = 23$ the number of opportunities is 253, the same as for the birthmate problem!

Even though the two problems are related through this idea of birthday opportunities, the relationship is not exact. Note that the formula obtained in the binomial approximation to the birthday problem, while an approximation of the birthday probability, is basically the same as the exact birthmate probability formula when the number of birthday opportunities are equated. The numeric probabilities for the two problems also indicate that while they can be related through this idea of birthday opportunities, the relationship is approximate. The probability of at least one match for the birthday problem for $n = 23$ is 0.5073 while for the birthmate problem the probability of at least one match for $n = 253$ is 0.5005.

Matching Up Your Sister

Suppose you go to a party and meet someone who has the same birthday as your sister. Many people would think that this is an unusual coincidence. But if there are seven people in your family then there need only be 33 people at the party to have a better than even chance of matching one birthday from your family. This type of coincidence is another variation of the birthday problem: What is the smallest number of people needed for a better than 50% chance that at least one pair has a matching birthday with one member of that pair in the first k individuals? In the above situation, your family serves as the first k individuals. (We'll use n to denote the total number of individuals at our party.)

To solve this we will again use the complement approach. For there to be no matches with one of the first k individuals, those first k individuals all must have different birthdays from each other and the other individuals still to be chosen. There are 365 choices for the first person's birthday. There are 364 choices for the second person's birthday since they must not match the first person's birthday. Continuing, we obtain $(365-k+1)$ choices for the kth person since they must not match any of the previously chosen $k-1$ days. Now the rest of the individuals cannot match any of the first k birthdays, but they can match each others' birthdays. This means for the remaining $n-k$ individuals each has $(365-k)$ choices for her birthday. The number of possible sets of birthdays for n people is 365^n. The probability of no person among the first k having a match with any of the others is

$$P(\text{no matches}) = \frac{365 \cdot 364 \cdots (365-k+1) \cdot (365-k)^{n-k}}{365^n}.$$

The probability of a match with one of the first k people is

$$P(\text{at least one match}) = 1 - \frac{365 \cdot 364 \cdots (365-k+1) \cdot (365-k)^{n-k}}{365^n}$$

The number of people involved must be at least 23, as seen in the birthday problem, or the probability will never exceed 50%. The following chart summarizes the smallest values of k and n that give a better than 50% probability of a match.

n	23	24	25	26	27	28	29	31	33	36	40	46	54	66	86	128	254
k	20	17	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

As before, we can use the idea of matching opportunities to get a good approximation to the solutions for this prob-

Illustration by Loel Barr

lem. Person 1 has $n - 1$ opportunities for a match, her birthday could match any of the remaining $n - 1$ people. Person 2 has $n - 2$ new opportunities for a match since his birthday could match any of the remaining $n - 2$ people (his birthday has already been compared to the first person). Continuing in this fashion, for person k, her birthday has $n - k$ new opportunities for a match. Summing the number of opportunities for the first k people gives $kn - k(k+1)/2 = k(n - (k+1)/2)$ as the total number of opportunities. Equating the number of opportunities to 253 and solving for n gives a formula that does a good job of finding (n, k) values that provide probabilities near 50% (the n values are rounded to the nearest integer). In fact in only one place ($k = 13$, $n = 26$) does it give a pair different from the above chart. But keep in mind this formula is only an approximation, just as the relationship between the birthday problem and the birthmate problem is approximate.

Matching Your Lottery Picks

Suppose you are working at your convenience mart and two consecutive customers come in and purchase lottery tickets. The tickets each have 6 numbers selected out of 45. You notice that one number is selected by both of the ticket buyers. You might think of this as a coincidence, but in fact, it turns out that there is a 60% chance that at least one number will be selected by both buyers. This is still another variation of the birthday problem: If each person is allowed to choose k distinct dates out of 365 days (6 distinct numbers on a lottery ticket out of 45), what is the smallest number of dates each person should choose to give a better than 50% chance of at least one match among the entire group? (Another way to think about it: how amazing a coincidence is it if my mother and your sister share a birthday?)

We would expect for small k similar probabilities to those for the original birthday problem. The problem here is that the dates are not all independent. Since each person chooses k distinct dates, those dates will be dependent and will not result in any matches within the group of k dates. Because of this we might expect that as k increases, more total birthdays will be needed to obtain a match.

To find the probability for this variation again consider the complement, the probability of no matches and construct the probability one person at a time. For the first person the probability of no matches is 1, since his chosen birthdays must be distinct. For the second person the probability of not matching any of the first person's k dates is $(365-k)/365 \cdot (365-k-1)/364 \cdots (365-2k+1)/(365-k+1)$. The numerators

$$(365 - k) \cdot (365 - k - 1) \cdots (365 - 2k + 1)$$

count the number of ways to choose the k dates that do not match the first person's k selections. The denominators are due to the k dates chosen by the second person being distinct. For the third person the probability of not matching any of the first two individual's dates is

$$\frac{365-2k}{365}\cdot\frac{365-2k-1}{364}\cdots\frac{365-3k+1}{365-k+1}.$$

This continues down until the nth person, whose probability of not matching any of the previously chosen dates is

$$\frac{365-(n-1)k}{365}\cdot\frac{365-(n-1)k-1}{364}\cdots\frac{365-nk+1}{365-k+1}.$$

The probabilities of these n people can then be multiplied to give the overall probability of no matches

$$P(\text{no matches})=\frac{365\cdot 364\cdots(365-nk+1)}{365^n\cdot 364^n\cdots(365-k+1)^n}$$

The probability of at least one match is then

$$P(\text{at least one match})=1-\frac{365\cdot 364\cdots(365-nk+1)}{365^n\cdot 364^n\cdots(365-k+1)^n}$$

The following table summarizes the smallest value of n that gives a probability of greater than 50% for various values of k.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
n	23	12	8	7	5	5	4	4	4	3	3	3	3	3	3	2

Notice that for each pair the product nk is in the neighborhood of 23 with the product generally straying further from 23 as k increases. Intuitively this makes sense since as k increases for a fixed total number of birthdays, nk, the number of opportunities for a match decreases because within each set of k birthdays chosen by an individual there cannot be a match. Therefore more total birthdays are needed for larger values of k.

We can also develop a birthday opportunities formula for this problem. In this case, Person 1, by himself, provides no opportunity for a match since his dates are distinct. Person 2 provides k^2 opportunities for a match since each of her k distinct dates could match any of the k dates of the first person. Person 3 provides $2k^2$ new opportunities for a match because each of his dates can match any of the $2k$ previously chosen dates. Continuing, the nth person provides $(n-1)k^2$ new opportunities for a match. Totaling the opportunities gives $k^2\cdot {}^{n(n-1)}\!/_2$. Equating this expression to 253 gives a formula to relate n and k values that provide probabilities near 50%. Solving this formula for k gives $k=\sqrt{{}^{506}\!/_{n(n-1)}}$. Here the formula works well for larger n ($7\le n\le 23$), but as n decreases to 6 or fewer the values of k given by the formula are less accurate. As already mentioned this is not too surprising because as k increases the independence assumption used in the birthmate problem is further violated by having a larger number of birthdays that cannot be matches of each other.

One of the amazing aspects of these problems is not so much the quantity of variations (only a handful have been considered here), but instead that so many of the variations are so interconnected to each other. But with mathematics, we expect that. Different legitimate approaches to the same problem should lead to the same answer. But it does make one wonder if 253 should be the number we remember when dealing with the birthday problem, rather than 23.

For Further Reading

The story of the double hijacking can be found in Hijacking Planes to Cuba: An Up-Dated Version of the Birthday Problem, *The American Statistician*, Vol. 24, No. 1 (1970), p. 41–44, by Ned Glick. The machine shop failure is described in Breakdowns and Birthdays, *Teaching Statistics*, Vol. 2(1980), p.15–18, by A. F. Bissell. The matching up your sister variation originated with Edmund A. Gehan in Note on the Birthday Problem, *The American Statistician*, Vol. 22 (1968). Neville Spencer devised the lottery pick variation in Celebrating the Birthday Problem, *The Mathematics Teacher*, Vol. 70, No. 4 (1977). Fred Mosteller, one of the great statisticians of the twentieth century, invented the birthmate problem and wrote about it and other fascinating probabilistic problems in Understanding the Birthday Problem, *The Mathematics Teacher*, Vol. 55 (1962), 322–325 and in *Fifty Challenging Problems in Probability*, Addison-Wesley, New York (1965).