Calculus meets the birthday problem

DALE K. HATHAWAY and JOSHUA BARKS

What is the likelihood $\mathrm{P}(n, T)$ that at least two people in a gathering of n people are born within a given time T of each other? In particular, for $T = 24$ hours what is the smallest n for which $\mathrm{P}(n, T)$ is at least 50%? The well-known classic version of the birthday problem asks for the smallest number of people needed to give a better than 50% chance of at least one birthday match with the assumptions that the birthdays are independently selected from a discrete uniform distribution over 365 days. Using the calculus tool of the limit, we refine two characterisations for $\mathrm{P}(n, T)$, and show that they give consistent results with each other and with the classic birthday problem as well. At first thought $\mathrm{P}(n,\ 24$ hours) should answer the classic birthday question. Yet consider this experiment: suppose Andrea was born at noon on June 1, then potential matching birth times (for the other $n - 1$ people) with her birth time extend from noon on May 31 to noon on June 2, a period of 48 hours! What this period suggests is that $\mathrm{P}(n,\ 24$ hours) exceeds the classic answer. But, by how much? Read on.

A little history

For the birthday problem, we assume that each year has exactly 365 days, and that birth dates follow a discrete uniform distribution. Following the work of Naus [1] and Abramson and Moser [2], let A_d be the event of at least two birthdays occurring within d adjacent calendar days, in which case $\mathrm{P}(A_1)$ is the classic birthday problem. Since it is much easier to solve the birthday problem by complements we will let A_d^c denote the complement of event A_d. Then

$$\mathrm{P}(A_1) = 1 - \mathrm{P}(A_1^c) = 1 - \frac{365!}{(365 - n)!\ 365^n} \tag{1}$$

and when $n = 23$ people, $\mathrm{P}(A_1) = 0.5073$. As Naus showed it is much easier to deal with birthdays on adjacent days if we know that there are no matching birthdays, so the probability of at least two birthdays within d calendar days of each other will be written in terms of no matching birthdays as

$$\mathrm{P}(A_d) = 1 - \mathrm{P}(A_d^c) = \mathrm{P}(A_d^c \mid A_1^c) \times \mathrm{P}(A_1^c). \tag{2}$$

To find the conditional probability $\mathrm{P}(A_d^c \mid A_1^c)$, the probability of no birthdays within d calendar days of each other given that there are no matching birthdays, note that if no birthdays are within d calendar days of each other, then there is no chance of a matching birthday and so each day is at most one person's birthday. Since the birthdays may be represented by a cycle of dates, the first birthday can be arbitrarily selected, and there are $n - 1$ birthdays remaining to be chosen from the remaining 364 days of the

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

year. The $n - 1$ remaining birthdays can be chosen in $\binom{364}{n-1}$ ways, which gives the denominator of $\mathrm{P}\left(A_d^c \mid A_1^c\right)$. To compute the numerator, consider the number of ways to choose n days from 365 with at least $d - 1$ days between any two birthdays. Think of the calendar days of the year being represented by a series of zeros and ones where a 1 represents a single birthday on that calendar day, and a 0 represents days with no birthdays. To ensure that the birthdays are at least d days apart consider each of the n birthdays to be immediately followed by $d - 1$ non-birthdays. This can be represented by n fixed subarrangements, each covering d consecutive days, all of the form

$$1\underbrace{0\ldots 0}_{d-1}.$$

For example, for $d = 4$, a snippet from a typical arrangement could look like

$$\ldots\ 0|1000|0|0|1000|1000|0|0|1000\ldots$$

where we use | to separate the groupings of either a birthday followed by 3 non-birthday dates or dates without birthdays.

The n fixed subarrangements, each of which occupies d consecutive days uses nd of the 364 days. The remaining $364 - nd$ days, which are all zeros, now need to be permuted with the n subarrangements. There are a total of $364 - nd + n$ items, of which $n - 1$ locations need to be chosen to fill with the $n - 1$ subarrangements since the first date has already been selected. There are $\binom{364 - n(d-1)}{n-1}$ ways to choose the locations. The probability of no birthdays within d calendar days of each other given no matching birthdays is then

$$\mathrm{P}\left(A_d^c \mid A_1^c\right) = \frac{\binom{364 - n(d-1)}{n-1}}{\binom{364}{n-1}} \tag{3}$$

which in turn by (2) gives Naus's formula,

$$\mathrm{P}\left(A_d\right) = \frac{(364 - n(d-1))!}{(365 - nd)! \times 365^{n-1}}. \tag{4}$$

From discrete to continuous time

Standing on the shoulders of Naus and others, like Spencer [3] who first suggested this variation, and Schwarz [4] who published a different solution, we now are ready to move from the discrete to the continuous by adapting (4) to an arbitrary time lapse T rather than a fixed period of d calendar days.

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

For the continuous variation considered here a generic version of this formula is needed, a version not tied to 365 discrete days (time periods) in a year. Let $P_{n,N}(A_d)$ denote the probability of at least two birthdays within d time periods of each other in a sample of n individuals with N possible time periods. Then

$$P_{n,N}(A_d) = 1 - \frac{((N-1) - n(d-1))!}{(N-nd)!\, N^{n-1}}$$

$$= 1 - \left(\frac{N-nd+1}{N}\right)\left(\frac{N-nd+2}{N}\right)\cdots\left(\frac{N-nd+n-1}{N}\right). \qquad (5)$$

Formula (5) with $N = 365$ and $d = 1$ gives the solution to the classic birthday problem. If we use hours as the incremental time unit, then $d = 24$ corresponds to 1 day, and $N = 365 \times 24 = 8760$ hours rather than 365 days. In the context of a 24-hour time period, N will always be 365 times d, or $d = \frac{N}{365}$. Substituting this expression for d into (5) gives

$$P_{n,N}(A_d) = 1 - \left(\frac{365-n}{365} + \frac{1}{N}\right)\left(\frac{365-n}{365} + \frac{2}{N}\right)\cdots\left(\frac{365-n}{365} + \frac{n-1}{N}\right). \qquad (6)$$

Now we lift our implied discrete restriction and interpret A_d as the event in which at least two of n people have birth times that are within d days of each other. Recall how the area under a curve is initially approximated in a calculus course by taking partitions with more subintervals which are smaller in width. We apply this approach here using discrete probability with shorter and shorter time units. In calculus we obtain the actual area under a curve by taking the limit, as the number of subintervals increases without bound, with their width approaching zero. Analogously here we take the limit as the number of time intervals increases without bound and their lengths go to zero. In taking this approach d represents our time period of interest, which may be any non-negative real number of days, but for simplicity when used as a subscript as in A_d, d will always be in days. Then N is the total number of incremental time units in a year.

Taking the limit as N approaches infinity gives the continuous probability, denoted by $P_{n,\infty}(A_1)$, that at least two of n people have birth times within one day of each other,

$$P_{n,\infty}(A_1) = 1 - \left(\frac{365-n}{365}\right)^{n-1}. \qquad (7)$$

Table 1 compares various n classic birthday probabilities (1) to continuous birthday probabilities (7).

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

n	5	10	15	16	17	20
$P_{n,365}(A_1)$	0.0271	0.1169	0.2529	0.2836	0.3150	0.4114
$P_{n,\infty}(A_1)$	0.0537	0.2212	0.4443	0.4895	0.5338	0.6572

n	22	23	25	30	40
$P_{n,365}(A_1)$	0.4757	0.5073	0.5687	0.8144	0.8912
$P_{n,\infty}(A_1)$	0.7290	0.7611	0.8178	0.9675	0.9892

TABLE 1: Probabilities that at least two of n people have birthdays within a day of each other

To generalise (7) to any non-negative real number t days, use the substitution $d = \frac{Nt}{365}$ on (5) to obtain

$$P_{n,\infty}(A_t) = 1 - \left[\frac{365 - nt}{365}\right]^{n-1}. \tag{8}$$

For example, for $t = \frac{1}{2}$ the smallest n for which $P_{n,\infty}(A_{1/2}) > 0.5$ is 23 which corresponds to the classic birthday problem $P_{n,365}(A_1)$. For $t = 2$ days, the least n for which $P_{n,\infty}(A_2) > 0.5$ is 12. For $t = \frac{1}{24} = 1$ hour, the least n for which $P_{n,\infty}(A_{1/24}) > 0.5$ is 79, and for $t = \frac{1}{1440}$ (one minute), n needs to be at least 604.

An alternative approach with continuous opportunities

An alternative approach to (8) for calculating the continuous probability of $P(A_d)$ is to use the opportunities approach for a match, first developed by Mosteller [5], and variants of which appear in [6, 7, 8]. The opportunities approach is based on the relationship between the birthmate problem, which is the probability of at least one match of a specific predetermined date, and the classic birthday problem. The number of people needed in the birthmate problem to ensure a probability of at least 50% is 253, the smallest n for which $\left(\frac{364}{365}\right)^n < 0.5$. Mosteller showed that the 23 people needed for the classic birthday problem corresponds to the 253 people needed for the birthmate problem through the number of opportunities for a match. There are $\binom{23}{2} = 253$ ways to select a pair of individuals from a set of 23 people. Any of these pairs could result in a birthday match between the two people involved in the pair. Each pairing can then be thought of as an opportunity for a birthday match. The probability that a particular pair does not produce a match is $1 - \frac{1}{365}$. Assuming the pairs are independent (they are actually not independent, they are only pairwise independent, but here that is good enough to give us an approximation), then to achieve a probability of at least 50% that there is at least one match among y possible pairs, we need the probability that none of the pairs yield a match to be less than 0.5, or

$$\left(1 - \frac{1}{365}\right)^y < 0.5, \tag{9}$$

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

also called the birthmate inequality. The probability of a birthday being within a day on a continuous scale means the matching birthday needs to occur within 24 hours either preceding or following the instant of the previous birthday. Consider the first person in each pair as fixing the birth time to be matched. The probability of a second birthday occurring within 24 hours of the fixed birth time is $\frac{2}{365}$ since it can precede or follow the first birthday. Replacing the $\frac{1}{365}$ with $\frac{2}{365}$ in the birthmate inequality above and solving for y gives

$$\left(1 - \frac{2}{365}\right)^y < 0.5$$

$$y > 126.152.$$

Therefore 127 opportunities, or pairs, are needed to give a better than 50% chance of birthdays within 24 hours of each other. Since $\binom{16}{2} = 120 < 127 < 136 = \binom{17}{2}$, at least 17 people are needed to give a better than 50% chance that at least two birthdays occur within 24 hours of each other. Even though these results are only approximate, they are consistent with previous results using the calculus area approach.

Table 2 shows the results of extending this approach to other values of d days.

d	1	2	3	4	5	6	7	8	9
y	127	63	42	32	25	21	18	16	14
n	17	12	10	9	8	7	7	7	6

TABLE 2: The number of opportunities (y) needed among n people so that P exceeds 50% for at least two birth times within d days.

The d and n values of Table 2 match the $d = t$ and n results for (8).

Final observations and conclusions

A simple relationship exists between the classic discrete and continuous birthday problem probabilities. For example we showed that $P_{n,365}(A_1)$ and $P_{n,\infty}(A_{1/2})$ both first exceed 50% for $n = 23$. Furthermore, with $N = 365 \times 12 = 4380$ (the number of 2-hour time units in a year) then $P_{n,4380}(A_{1/12})$ and $P_{n,\infty}(A_{1/24})$ both first exceed 50% for $n = 79$. Similarly, with $N = 365 \times 24 = 8760$ (the number of hours in a year) then $P_{n,8760}(A_{1/24})$ and $P_{n,\infty}(A_{1/48})$ both first exceed 50% for $n = 111$. The continuous case corresponds to a version of the classic birthday problem with a time period of half the size. For example, a person born at 12:01 p.m. on January 13th is within an hour of everyone born between 11:01 a.m. and 1:01 p.m. on January 13th, a time period of two hours. This can be explained heuristically by considering the birth time to occur in the middle

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

of the time period. In the classic birthday problem consider all birthdays to occur at noon so that 12 hours on either side of the birthday gives a calendar day. A 12-hour continuous time period requires 23 people for the birthday probability. The $d = 2$ discrete case can be thought of as the first birthday again being placed at noon of the day on which it occurs. Including 36 hours on either side of the noon birthday would give birthdays within the same or consecutive calender days. The discrete value of $n = 14$ for $d = 2$ again matches the continuous value for an interval of 36 hours or 1.5 days. In the context of birthdays within $d \geqslant 1$ days of each other the discrete case with k hours corresponds to the continuous case with $k - 12$ hours.

Some authors approach the birthday problem by disallowing matches across the year end, so that December 31 is not considered next to January 1. In this restricted context, the reader may verify that (7) becomes

$$P_{n,\infty}(A_1) = 1 - \left[\frac{365 - n}{365}\right]^n \tag{10}$$

with similar adjustments in (6) and (8).

In conclusion if we view the birthdays as continuous on the interval [0, 365] and a day as an arbitrary 24-hour period of time, then it requires fewer people to give a better than 50% chance of a birthday match of at least two within 24 hours of each other. Of course in practice this is more difficult to demonstrate using actual birthdays, since while everyone knows their birthday, not everyone knows their birth time. But the next time you demonstrate the birthday problem, if you do not get a match but get birthdays on consecutive days, you might want to ask the people involved if they know the time of day that they were born. You might have a continuous birthday match.

Acknowledgements

The authors would like to thank the editor and an anonymous referee for their helpful suggestions, as well as Andrew Simoson who helped improve the readability of the article. The work for this article was done while the second author was an undergraduate supervised by the first author and supported by the Pence-Boyce Undergraduate Summer Research Program at Olivet Nazarene University.

References

1. Joseph Naus, An extension of the birthday problem, *The American Statistician* **22** 1 (1968) pp. 27-29.
2. Morton Abramson and W. O. J. Moser, More birthday surprises, *Amer. Math. Monthly* **77** 8 (1970) pp. 856-858.
3. Neville Spencer, Celebrating the birthday problem, *Mathematics Teacher* **70** 4 (1977) pp. 348-353.

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press

4. Wolf Schwarz, Comparing continuous and discrete birthday coincidences: 'same-day' versus 'within 24 hours', *The American Statistician* **64** 1 (2010) pp. 34-36.
5. Frederick Mosteller, Understanding the birthday problem, *Mathematics Teacher* **55** 5 (1962) pp. 322-325.
6. Dale K. Hathaway, A bonanza of birthday bewilderments, *Math Horizons* **7** (2001) pp. 13-16.
7. Lawrence M. Lesser, Exploring the birthday problem with spreadsheets, *Mathematics Teacher* **92** 5 (1999) pp. 407-411.
8. Alan H. Rapoport, Birthday problems – a search for elementary solutions, *Math. Gaz*. **82** (March 1998) pp. 111-114.

doi:10.1017/mag.2016.51

DALE K. HATHAWAY
Olivet Nazarene University, One University Avenue, Bourbonnais, IL 60914, USA
e-mail: *hathaway@olivet.edu*
JOSHUA BARKS
Nestle USA, 2501 Beich Rd., Bloomington, IL 61705, USA
e-mail: *barksj11@yahoo.com*

Farewell

Many individuals help with the *Gazette* in many different ways. Sadly, one of these is stepping down. Richard Kirby has been checking the *Gazette Index* for a long time. He has done this meticulously for 14 years despite difficulties. Sadly, his personal circumstances are becoming increasingly worse and he cannot continue. Thank you Richard, you have the eyes of an eagle.

This creates a vacancy which we would like to fill. The Index is constructed by Bill Richardson soon after the November issue is completed and then sent for checking. Errors at this stage might include incorrect alphabetical order, omitting a book author and such like. There is no financial reward, just our thanks and a feeling of satisfaction. If you are interested please contact me.

GERRY LEVERSHA
e-mail: *g.leversha@btinternet.com*

https://doi.org/10.1017/mag.2016.51 Published online by Cambridge University Press