A Closer Look at the Notorious Birthday Coincidences

Ruma Falk
The Hebrew University of Jerusalem, Israel
e-mail: rfalk@cc.huji.ac.il

Summary The article asks about the minimal number of persons required for achieving a probability 1/2 that **a**. At least two share a birthday, **b**. At least one shares the reader's birthday. A basic question about the necessary number of checks underlies both problems.

Keywords: Teaching; Birthdays; Coincidences; Counter-intuitive results; Under/Overestimation; Combinatorial growth.

INTRODUCTION

It stands to reason that many a reader would react to the article's title by 'I know this problem already'. Indeed, most of us have been familiar with the birthday teaser way back. It represents a classic epitome of a counter-intuitive probability result.

Much has been written about random matches of birthdays (see Nickerson 2004, pp. 87–88 for a concise review) and about the surprisingness (or disbelief) following the finding that for 23 people, the probability that at least two have the same birthday exceeds 1/2 (e.g. Weaver 1963, pp. 132–135). These writings contain meticulous proofs, the numerical consequences of which were considered 'astounding' by Feller (1957, p. 32). Many authors developed generalizations and ramifications of the above standard problem (e.g. Diaconis and Mosteller 1989, Matthews and Stones 1998 and McKinney 1966). Seemingly, there is no point in additional delving into the question of birthday coincidences. Yet, I chose to revisit this issue so as to highlight one coincidence of numbers that had been almost completely overlooked in the cogent expositions that I had surveyed.

Two fundamental problems of shared birthdays will be presented with their solutions. We'll try to understand the reasons for the prevalence of people's erroneous intuitive estimates. The lesson drawn from the combination of the two analyses could be utilized didactically to settle any lingering doubts that students might harbour with respect to the birthday puzzle.

We assume equal chances of being born in all 365 days of the year (ignoring leap years and multiple births) and independence between dates of different births. A birthday is defined by the day in a month, ignoring the year of birth.

TWO PROBLEMS

Following the lead of Michalewicz and Michalewicz (2008, pp. 142–145) and Mosteller (1965, Problems 31 and 32), let's try to find the numbers sought out in two puzzles:

Problem **a**. Find the smallest number of people that should be in a room to make a bet with a better than 50% chance of winning that there are among them at least two with the same birthday.
Problem **b**. Find the smallest number of other people that should be in a room with you to make a bet with a better than 50% chance of winning that at least one of them has the same birthday as yours.

THE TWO SOLUTIONS

An expedient way to find the probability of the event that at least two persons share the same birthday, or that at least one shares your birthday, is to compute the probability of the complementary event that they all have distinct birthdays, or that none shares your birthday.

Without loss of generality, we can number the people in both problems in any arbitrary way we wish.

In case **a**, the probability that number 2's birthday differs from that of number 1 is 364/365, the probability that number 3's birthday differs from the two first ones is 363/365 and the probability that number k's birthday differs from those of 1, 2, …, $k-1$ is $(365-k+1)/365$. All k birthdays differ from each other with a probability of

$$364 \times 363 \times \ldots \times (365 - k + 1)/365^{k-1}$$

Hence, the probability of at least one shared birthday among k people is

$$1 - \frac{364 \times 363 \times \cdots \times (365 - k + 1)}{365^{k-1}} \quad (1)$$

Analogously, in case **b**, the probability that number 1's birthday differs from yours is 364/365, and so are the probabilities of numbers 2, 3, …, k. So the probability that k independent birthdays differ from yours is $(364/365)^k$ and that at least one birthday among the k matches yours is

$$1 - \left(\frac{364}{365}\right)^k \quad (2)$$

The minimal integer k for which the probability of interest surpasses 1/2 is:
23 for formula (1) in Problem **a**;
253 for formula (2) in Problem **b**.

In reality, birthdays are not exactly equally likely in all 365 days of the year. Yet, Berresford (1980) computed the probability of a shared birthday, based on empirical birth rates (from New York State in 1977) that differed from homogeneity across the year, and found that answer **a** was surprisingly robust: 23 people were required for raising the probability above 1/2. He showed that in a group of people, the probability of a shared birthday is **least** for the uniform distribution; therefore, regardless of the actual distribution, 23 people suffice for making a shared birthday more probable than not.

Likewise, Matthews and Stones (1998) maintained that real-life deviations from the theoretical uniformity of birthday rates over the year tend to favour the existence of more coincidences. They also checked the birth dates of people in 10 groups, each comprising 23 men on the football pitch (11 in each of the two teams plus the referee). They found impressive agreement between the predictions based on the slightly simplistic mathematical assumptions and the real observed numbers of coincidences in the 10 football fixtures (Table 2). See also Petocz and Sowey (2006) on 'common birthday among British Prime Ministers' (p. 63).

COMMON GUESSES

The number suggested offhandedly by my respondents (students, colleagues and friends) as an answer to both problems has usually been **183** (half of 365 rounded upwards) or close to it. The same was reported by Michalewicz and Michalewicz (2008), Mosteller (1962), and in the expositions of Plous (1993, pp. 150–151) and Rosenhouse (2009, pp. 5–7). These respondents probably felt that, given 365 opportunities, there is no reason to crowd together unless the number of candidates reaches about half the available slots. The pigeonhole principle assures us that if n items are put into m pigeonholes, with $n > m$, then at least one pigeonhole must contain more than one item. The birthday problems require only half that certainty, so one reckons that the number of items should be no less than half the number of pigeonholes.

Intuitive estimates thus miss the mark. In Problem **a**, they overestimate the correct answer, and in Problem **b**, they underestimate it.

WHY LAY ESTIMATES ARE WIDE OF THE MARK

Apparently, besides relying on 'half the pigeonhole principle', many assume in Problem **a** a self-centred perspective (Mosteller 1962; Rosenhouse 2009, p. 6) and think about how many people are required for having 50–50 chances of at least one match with their own birthday (disregarding possible matches between pairs of other persons). In that case, they are right to propose a number much greater than 23 (and still not big enough). But they answer the wrong question. Those who take into account possible shared birthdays of others appraise the number of pairs of others far below the number of combinations of two elements out of a given finite set. In Problem **b**, which does focus on the solvers' own birthdays, their image of 183 persons occupying about half the dates of a year is flawed because they are apparently oblivious to the substantial number of possibilities of two or more persons who might have common birthdays and make the 183 cover much less than half the available dates.

In both cases, the naïve assessments underestimate grossly the number of pairs (triplets etc.) whose birthdays might coincide. Consistent underestimation of exponential growth had been reported by Wagenaar and his associates (e.g. Wagenaar and Timmers 1979) and that of multiplicative or factorial growth by Tversky and

Kahneman (1974). In the birthdays case, considerable underestimation of combinatorial growth accounts for an exaggerated image of the spread of 183 birthdays over the year and for falling short of the number of pairs that can be formed in a group of 23 persons. That number—$C(23, 2) = \mathbf{253}$—is much greater than the 22 possible pairs of the solver and other members of the group, and even than the lay assessments of the total number of possible pairs.

A CORE INSIGHT

The reappearance of the number 253—which had solved Problem **b**—in the context of Problem **a** struck me first as 'just a coincidence' that may be dismissed. On second thoughts, however, I found it enlightening. The repetition of this number can, in fact, be considered a metacoincidence, namely a coincidence of a higher order. And it is not a fluke of chance. The key number **253** answers a basic question that is explicit in Problem **b** and rather implicit in Problem **a**.

> The **basic question**, the answer to which underlies the solution of both problems, is
>
> How many independent checks of pairs of birthdays —each with success probability 1/365—are required for obtaining 50% chances of at least one success?

It makes no difference whether all the checks involve the same focal person paired with all the others (as in case **b**) or whether each of the people is included several times but is always paired with somebody else (as in case **a**). As long as all the (unordered) pairs are different, the outcomes of the checks are all independent of each other.

The crux of the matter is that what counts in both problems is the key number of pairs that have to be checked for a match or a lack thereof. Notably, this number is the same in the two problems. Michalewicz and Michalewicz (2008) presented both problems and got the answer 253 to Problem **b** but failed to connect it to the number of checks that should be performed in Problem **a**. And Rosenhouse (2009) rightly maintained, in solving Problem **a**, that 'With 23 people there are 253 pairs, which means 253 chances of getting a match' (p. 7). This should only have been rounded off with the finding that 253 solves Problem **b**. After completing my analysis, I was delighted to find that Mosteller, in his 1962 article and in Problem 33 of his 1965 book, related Problems **a** and **b** to each other and showed that for exceeding the probability 1/2, both require the same number of checks, which is the number of opportunities for like birthdays. And for any target probability, not just 1/2, the number r of pairs to check in Problem **a** and the number n of people to check in Problem **b** satisfy $n \approx r(r-1)/2$.

PRAGMATIC CONCLUSIONS

The number 253, which had been given above without proof as the solution of Problem **b**, is found by solving for n (the exponent in computing the complementary probability for Problem **b** and the **basic question**) in the equation

$$\left(\frac{364}{365}\right)^n = \frac{1}{2} \qquad (3)$$

Taking logarithms of both sides easily yields 253 as the closest integer answer. Note that the exact non-integer solution is $n = 252.652$; hence, also other integer results are approximate (see Mosteller's 1962 table on p. 325).

Although 253 solves the **basic question** of both problems, it does not give us the size of the set sought in Problem **a**. However, this can easily be derived from 253. Let r be the unknown answer to Problem **a**, clearly, 253 equals $C(r, 2) = r(r-1)/2$, and r can be extracted by solving the quadratic equation

$$r(r-1)/2 = 253$$

the positive root of which is 23. Note that it is technically easier to obtain the answer 23 this way than to equate the subtrahend of formula (1) (or an equivalent form thereof) to 1/2, as performed in the presentations of the problem in the literature, where the computation involves approximations (e.g. Diaconis and Mosteller 1989, p. 857 and Feller (1957, p. 32)).

The advantage of the traditional solution of Problem **a** by means of formula (1) lies in its conceptual clarity. It is easily comprehensible but hard to manipulate practically. Solving via formula (3) is pragmatically preferable. It might obviate the need for the traditionally cumbersome solutions. More than that, this method serves to highlight the pivotal role of the number of pairs that have to be checked in solving different varieties of birthday problems.

Acknowledgement

I thank Raphael Falk for both conceptual and technical help.

References

Berresford, G. C. (1980). The uniformity assumption in the birthday problem. *Mathematics Magazine*, **53**(5), 286-288.
Diaconis, P. and Mosteller, F. (1989). Methods for studying coincidences. *Journal of the American Statistical Association*, **84**(No. 408, Application and Case Studies), 853-861.
Feller, W. (1957). An Introduction to Probability Theory and Its Applications (2nd edn, Vol. **1**). New York: Wiley.
Matthews, R. and Stones, F. (1998). Coincidences: The truth is out there. *Teaching Statistics*, **20**(1), 17-19.
McKinney, E. H. (1966). Generalized birthday problem. *The American Mathematical Monthly*, **73**(4), 385-387.
Michalewicz, Z. and Michalewicz, M. (2008). Puzzle-Based Learning: An Introduction to Critical Thinking, Mathematics, and Problem Solving. Melbourne Victoria, Australia: Hybrid Publishers.
Mosteller, F. (1962). Understanding the birthday problem. *Mathematics Teacher*, **55**(5), 322-325.
Mosteller, F. (1965). Fifty Challenging Problems in Probability with Solutions. Reading, MA: Addison-Wesley.
Nickerson, R. S. (2004). Cognition and Chance: The Psychology of Probabilistic Reasoning. Mahwah, NJ: Lawrence Erlbaum.
Petocz, P. and Sowey, E. (2006). Statistical diversions. *Teaching Statistics*, **28**(2), 61–64.
Plous, S. (1993). The Psychology of Judgment and Decision Making. New York: McGraw-Hill.
Rosenhouse, J. (2009). The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser. Oxford: Oxford University Press.
Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124-1131.
Wagenaar, W. A. and Timmers, H. (1979). The pond-and-duckweed problem: Three experiments on the misperception of exponential growth. *Acta Psychologica*, **43**, 239-251.
Weaver, W. (1963). Lady Luck: The Theory of Probability. New York: Dover.

Postscript Note

Thanks are due to Raymond Nickerson who had read the draft of my article and sent me his chart of the probability of at least one birthday match, plotted as a function of the number of people in the group, for the two birthday problems. After completing the procedure for publication of my paper, I obtained the Spring 2013 issue of *Teaching Statistics* (volume 35, no. 1), where the same lovely figure appears in Matthew Russell's article on page 27.

Ruma Falk

A shortened version of this note has been offered as a companion piece as it includes visuals that usefully help explain these problems.

Paul Hewson

On why it is easier to find two People with the same Birthday than one other Person with one's own

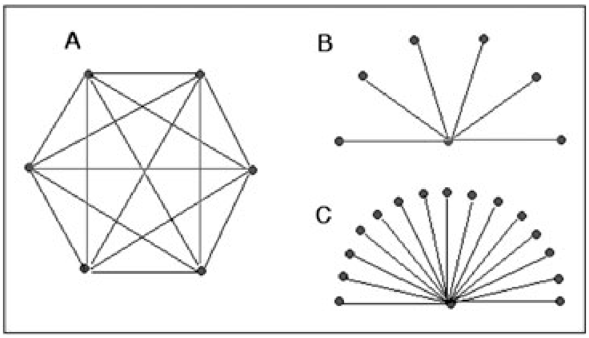
Raymond S. Nickerson
Tufts University

It is well known by students of probability theory that the probability that at least two of the people in a random group of 23 will have the same birthday is slightly more than .5. This is of interest because when asked how many random people it would take to make the probability that at least two would have the same birthday, the estimate typically given is much larger than 23; according to Mosteller (1962), a common answer is about 183, approximately half the number of days in a year. Less well known is that the number of random people required to make the probability of finding at least one with the same birthday as one's own to be slightly more than 0.5 is 253 (Feller, 1957). In what follows, I will use the terminology of Mosteller (1962), who referred to the probability of two random people having the same birthday as the *birthday* problem and to the probability of a random person having the same birthday as one's own (or a specific other) as the *birthmate* problem.

It is intuitively obvious that the probability that at least two people from a group of n random people will have the same birthday increases with n. Similarly, it is clear that the probability that at least one person from a group of n random people will have the same birthday as one's own (i.e. will have a birthday on a specified date) also increases with n, albeit at a greatly different rate (Russell, 2013). Mosteller (1962) noted that the numbers 23 and 253 are linked by a combinatorial equation: 253 is the number of combinations of 23 things taken 2 at a time, i.e. $\binom{23}{2} = \frac{(23)(22)}{2} = 253$. The question naturally arises as to whether this is an insignificant happenstance or a clue to something interesting.

Recently, Falk (2013) pointed out that this is not a chance coincidence, but rather, it reveals a commonality between the two problems, namely that what counts in both cases is 'the number of pairs that must be checked for a match or a lack thereof'. It follows that for any value of $p<1$, either value of n can be inferred from the other. Thus, e.g. for $p=0.25$, if one knows the birthday number to be 15, as it is, then one can infer that the birthmate number is $\binom{15}{2} = \frac{(15)(14)}{2} = 105$. Or, conversely, if one knows the birthmate number is 105, then one can infer by taking the positive root of the quadratic $[x(x-1)]/2 = 105$ that the birth rate number is 15.

Falk's insightful noting that the birthday and birthmate problems require that the same number of pairs be checked for matches shines a very helpful light on the problems. In particular, it reveals a causal connection between the numbers that describe the sample sizes required to attain specified probabilities of finding matches in the two cases. The situation may be further illuminated by the following figure.



Drawing A represents the birthday problem, showing the 15 pairs (i.e. the 15 line segments that connect pairs of the six points) that must be checked to see if there are any common birthdays among six people. Drawings B and C represent the birthmate problem, and show, respectively, all the ways in which the birthdays of 6 and 15 people can be compared with the birthday of a single person. The critical thing to see is that for the number of comparisons that must be made to be equal in the birthday and birthmate problems, the number of people must be much larger in the birthmate case than in the birthday case. Given a group of n people, the number of pairings that must be considered in the birthday case, in which everyone's birthday must be compared with that of everyone else, is the number of ways in which *n* things can be combined two at a time, i.e. $\binom{n}{2} = \frac{(n)(n-1)}{2}$. The number of pairings that must be considered in the birthmate case, in which everyone's birthday must be compared with that of a specific other, is simply the number of people in the group, let us call it *m*. So the required number of pairings in the one case will equal that in the other when $m = \binom{n}{2}$. Incidentally, the probability of finding two common birthdays among six random people is 0.04, which is the same as the probability of finding a birthmate among 15 random people.

References

Falk, R. (2013). A closer look at the notorious birthday coincidences. *Teaching Statistics*.

Feller, W. (1957). An Introduction to Probability Theory and its Applications (2nd edn, Vol. **1**). New York: Wiley. (Originally published in 1950).

Mosteller, F. (1962). Understanding the birthday problem. *The Mathematics Teacher*, **55**(5), 322-325.

Russell, M. (2013). Pigeons, Facebook and the birthday problem. *Teaching Statistics*, **39**, 26-28.

AUTHOR'S NOTE

This note grew out of correspondence with Ruma Falk, whom I thank for encouragement to write it and for helpful comments on a draft of it.