Birthday problem

In probability theory, the **birthday problem** asks for the probability that, in a set of n randomly chosen people, at least two will share the same birthday. The **birthday paradox** is the counterintuitive fact that only 23 people are needed for that probability to exceed 50%.

The birthday paradox is a veridical paradox: it seems wrong at first glance but is, in fact, true. While it may seem surprising that only 23 individuals are required to reach a 50% probability of a shared birthday (the number being less than 1/15th of the number of days in a year), this result is made more intuitive by considering that the birthday comparisons will be made between every possible pair of individuals. With 23 individuals, there are $\frac{23 \times 22}{2} = 253$ pairs to consider.

The computed probability of at least two people sharing the same birthday versus the number of people

Real-world applications for the birthday problem include a cryptographic attack called the birthday attack, which uses this probabilistic model to reduce the complexity of finding a collision for a hash function, as well as calculating the approximate risk of a hash collision existing within the hashes of a given size of population.

The problem is generally attributed to Harold Davenport in about 1927, though he did not publish it at the time. Davenport did not claim to be its discoverer "because he could not believe that it had not been stated earlier".[1][2] The first publication of a version of the birthday problem was by Richard von Mises in 1939.[3]

Calculating the probability

Consider the event A that a group of k people does not have any repeated birthdays, and let the complementary event B be that of a group of k people contains at least two people who share a birthday. Then the probabilities $P(A)$ and $P(B)$ of the two events are related by the equation $P(B) = 1 - P(A)$. The probability $P(A)$ can be computed using the perspective of permutations, as follows. Let V_{nr} be the total number of ways that k people can have distinct birthdays, and let V_t be the total number of ways k people can have birthdays arranged, including possibly repeated birthdays. The probability $P(A)$ is the ratio of these two quantities, V_{nr} divided by V_t. When $k = 23$, the two counts are given by

$$V_{nr} = 365 \cdot 364 \cdot 363 \cdots (365 - 22) = \frac{365!}{(365 - 23)!}$$

and

$$V_t = 365 \cdot 365 \cdot 365 \cdots 365 = 365^{23},$$

their ratio is $P(A) = \frac{V_{nr}}{V_t} \approx 0.492703$, and so

$$P(B) = 1 - P(A) \approx 1 - 0.492703 = 0.507297 \quad (50.7297\%).$$

Another way the birthday problem can be solved is by asking for an approximate probability that in a group of n people at least two have the same birthday. For simplicity, leap years, twins, selection bias, and seasonal and weekly variations in birth rates[4] are generally disregarded, and instead it is assumed that there are 365 possible birthdays, and that each person's birthday is equally likely to be any of these days, independent of the other people in the group.

For independent birthdays, a uniform distribution of birthdays minimizes the probability of two people in a group having the same birthday. Any unevenness increases the likelihood of two people sharing a birthday.[5][6] However real-world birthdays are not sufficiently uneven to make much change: the real-world group size necessary to have a greater than 50% chance of a shared birthday is 23, as in the theoretical uniform distribution.[7]

The goal is to compute $P(B)$, the probability that at least two people in the room have the same birthday. However, it is simpler to calculate $P(A')$, the probability that no two people in the room have the same birthday. Then, because B and A' are the only two possibilities and are also mutually exclusive, $P(B) = 1 - P(A')$.

Here is the calculation of $P(B)$ for 23 people. Let the 23 people be numbered 1 to 23. The event that all 23 people have different birthdays is the same as the event that person 2 does not have the same birthday as person 1, and that person 3 does not have the same birthday as either person 1 or person 2, and so on, and finally that person 23 does not have the same birthday as any of persons 1 through 22. Let these events be called Event 2, Event 3, and so on. Event 1 is the event of person 1 having a birthday, which occurs with probability 1. This conjunction of events may be computed using conditional probability: the probability of Event 2 is $\frac{364}{365}$, as person 2 may have any birthday other than the birthday of person 1. Similarly, the probability of Event 3 given that Event 2

occurred is $\frac{363}{365}$, as person 3 may have any of the birthdays not already taken by persons 1 and 2. This continues until finally the probability of Event 23 given that all preceding events occurred is $\frac{343}{365}$. Finally, the principle of conditional probability implies that $P(A')$ is equal to the product of these individual probabilities:

$$P(A') = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \cdots \times \frac{343}{365} \quad (1)$$

The terms of equation (**1**) can be collected to arrive at:

$$P(A') = \left(\frac{1}{365}\right)^{23} \times (365 \times 364 \times 363 \times \cdots \times 343) \quad (2)$$

Evaluating equation (**2**) gives $P(A') \approx 0.492703$

Therefore, $P(B) \approx 1 - 0.492703 = 0.507297$ (50.7297%).

This process can be generalized to a group of n people, where $p(n)$ is the probability of at least two of the n people sharing a birthday. It is easier to first calculate the probability $\bar{p}(n)$ that all n birthdays are *different*. According to the pigeonhole principle, $\bar{p}(n)$ is zero when $n > 365$. When $n \leq 365$:

$$\begin{aligned}\bar{p}(n) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{n-1}{365}\right) \\ &= \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n} \\ &= \frac{365!}{365^n (365-n)!} = \frac{n! \cdot \binom{365}{n}}{365^n} = \frac{{}_{365}P_n}{365^n}\end{aligned}$$

where $!$ is the factorial operator, $\binom{365}{n}$ is the binomial coefficient and ${}_kP_r$ denotes permutation.

The equation expresses the fact that the first person has no one to share a birthday, the second person cannot have the same birthday as the first $\left(\frac{364}{365}\right)$, the third cannot have the same birthday as either of the first two $\left(\frac{363}{365}\right)$, and in general the nth birthday cannot be the same as any of the $n - 1$ preceding birthdays.

The event of at least two of the n persons having the same birthday is complementary to all n birthdays being different. Therefore, its probability $p(n)$ is

$$p(n) = 1 - \bar{p}(n).$$

The following table shows the probability for some other values of n (for this table, the existence of leap years is ignored, and each birthday is assumed to be equally likely):

The probability that no two people share a birthday in a group of n people. Note that the vertical scale is logarithmic (each step down is 10^{20} times less likely).

n	$p(n)$
1	0.0%
5	2.7%
10	11.7%
20	41.1%
23	50.7%
30	70.6%
40	89.1%
50	97.0%
60	99.4%
70	99.9%
75	99.97%
100	99.999 97%
200	$(100 - 2 \times 10^{-28})\%$
300	$(100 - 6 \times 10^{-80})\%$
350	$(100 - 3 \times 10^{-129})\%$
365	$(100 - 1.45 \times 10^{-155})\%$
≥ 366	100%

Approximations

The Taylor series expansion of the exponential function (the constant $e \approx 2.718\ 281\ 828$)

provides a first-order approximation for e^x for :

To apply this approximation to the first expression derived for $\bar{p}(n)$, set $x = -\frac{a}{365}$. Thus,

Then, replace a with non-negative integers for each term in the formula of $\bar{p}(n)$ until $a = n - 1$, for example, when $a = 1$,

Graphs showing the approximate probabilities of at least two people sharing a birthday (red) and its complementary event (blue)

The first expression derived for $\bar{p}(n)$ can be approximated as

Therefore,

An even coarser approximation is given by

which, as the graph illustrates, is still fairly accurate.

A graph showing the accuracy of the approximation $1 - e^{-n^2/730}$ (red)

According to the approximation, the same approach can be applied to any number of "people" and "days". If rather than 365 days there are d, if there are n persons, and if $n \ll d$, then using the same approach as above we achieve the result that if $p(n, d)$ is the probability that at least two out of n people share the same birthday from a set of d available days, then:

Simple exponentiation

The probability of any two people not having the same birthday is $\frac{364}{365}$. In a room containing n people, there are $\binom{n}{2} = \frac{n(n-1)}{2}$ pairs of people, i.e. $\binom{n}{2}$ events. The probability of no two people sharing the same birthday can be approximated by assuming that these events are independent and hence by multiplying their probability together. Being independent would be equivalent to picking with replacement, any pair of people in the world, not just in a room. In short $\frac{364}{365}$ can be multiplied by itself $\binom{n}{2}$ times, which gives us

Since this is the probability of no one having the same birthday, then the probability of someone sharing a birthday is

And for the group of 23 people, the probability of sharing is

Poisson approximation

Applying the Poisson approximation for the binomial on the group of 23 people,

so

The result is over 50% as previous descriptions. This approximation is the same as the one above based on the Taylor expansion that uses $e^x \approx 1 + x$.

Square approximation

A good rule of thumb which can be used for mental calculation is the relation

which can also be written as

which works well for probabilities less than or equal to $\frac{1}{2}$. In these equations, d is the number of days in a year.

For instance, to estimate the number of people required for a $\frac{1}{2}$ chance of a shared birthday, we get

Which is not too far from the correct answer of 23.

Approximation of number of people

This can also be approximated using the following formula for the *number* of people necessary to have at least a $\frac{1}{2}$ chance of matching:

This is a result of the good approximation that an event with $\frac{1}{k}$ probability will have a $\frac{1}{2}$ chance of occurring at least once if it is repeated $k \ln 2$ times.[8]

Probability table

length of hex string	no. of bits (b)	hash space size (2^b)	Number of hashed elements such that probability of at least one hash collision ≥ p									
			$p = 10^{-18}$	$p = 10^{-15}$	$p = 10^{-12}$	$p = 10^{-9}$	$p = 10^{-6}$	$p = 0.001$	$p = 0.01$	$p = 0.25$	$p = 0.50$	$p = 0.75$
8	32	4.3×10^9	2	2	2	2.9	93	2.9×10^3	9.3×10^3	5.0×10^4	7.7×10^4	1.1×10^5
(10)	(40)	(1.1×10^{12})	2	2	2	47	1.5×10^3	4.7×10^4	1.5×10^5	8.0×10^5	1.2×10^6	1.7×10^6
(12)	(48)	(2.8×10^{14})	2	2	24	7.5×10^2	2.4×10^4	7.5×10^5	2.4×10^6	1.3×10^7	2.0×10^7	2.8×10^7
16	64	1.8×10^{19}	6.1	1.9×10^2	6.1×10^3	1.9×10^5	6.1×10^6	1.9×10^8	6.1×10^8	3.3×10^9	5.1×10^9	7.2×10^9
(24)	(96)	(7.9×10^{28})	4.0×10^5	1.3×10^7	4.0×10^8	1.3×10^{10}	4.0×10^{11}	1.3×10^{13}	4.0×10^{13}	2.1×10^{14}	3.3×10^{14}	4.7×10^{14}
32	128	3.4×10^{38}	2.6×10^{10}	8.2×10^{11}	2.6×10^{13}	8.2×10^{14}	2.6×10^{16}	8.3×10^{17}	2.6×10^{18}	1.4×10^{19}	2.2×10^{19}	3.1×10^{19}
(48)	(192)	(6.3×10^{57})	1.1×10^{20}	3.5×10^{21}	1.1×10^{23}	3.5×10^{24}	1.1×10^{26}	3.5×10^{27}	1.1×10^{28}	6.0×10^{28}	9.3×10^{28}	1.3×10^{29}
64	256	1.2×10^{77}	4.8×10^{29}	1.5×10^{31}	4.8×10^{32}	1.5×10^{34}	4.8×10^{35}	1.5×10^{37}	4.8×10^{37}	2.6×10^{38}	4.0×10^{38}	5.7×10^{38}
(96)	(384)	(3.9×10^{115})	8.9×10^{48}	2.8×10^{50}	8.9×10^{51}	2.8×10^{53}	8.9×10^{54}	2.8×10^{56}	8.9×10^{56}	4.8×10^{57}	7.4×10^{57}	1.0×10^{58}
128	512	1.3×10^{154}	1.6×10^{68}	5.2×10^{69}	1.6×10^{71}	5.2×10^{72}	1.6×10^{74}	5.2×10^{75}	1.6×10^{76}	8.8×10^{76}	1.4×10^{77}	1.9×10^{77}

The lighter fields in this table show the number of hashes needed to achieve the given probability of collision (column) given a hash space of a certain size in bits (row). Using the birthday analogy: the "hash space size" resembles the "available days", the "probability of collision" resembles the "probability of shared birthday", and the "required number of hashed elements" resembles the "required number of people in a group". One could also use this chart to determine the minimum hash size required (given upper bounds on the hashes and probability of error), or the probability of collision (for fixed number of hashes and probability of error).

For comparison, 10^{-18} to 10^{-15} is the uncorrectable bit error rate of a typical hard disk.[9] In theory, 128-bit hash functions, such as MD5, should stay within that range until about 8.2×10^{11} documents, even if its possible outputs are many more.

An upper bound on the probability and a lower bound on the number of people

The argument below is adapted from an argument of Paul Halmos.[nb 1]

As stated above, the probability that no two birthdays coincide is

As in earlier paragraphs, interest lies in the smallest n such that $p(n) > \frac{1}{2}$; or equivalently, the smallest n such that $\bar{p}(n) < \frac{1}{2}$.

Using the inequality $1 - x < e^{-x}$ in the above expression we replace $1 - \frac{k}{365}$ with $e^{-k/365}$. This yields

Therefore, the expression above is not only an approximation, but also an upper bound of $\bar{p}(n)$. The inequality

implies $\overline{p}(n) < \frac{1}{2}$. Solving for n gives

Now, $730 \ln 2$ is approximately 505.997, which is barely below 506, the value of $n^2 - n$ attained when $n = 23$. Therefore, 23 people suffice. Incidentally, solving $n^2 - n = 730 \ln 2$ for n gives the approximate formula of Frank H. Mathis cited above.

This derivation only shows that *at most* 23 people are needed to ensure the chances of a birthday match are at least even; it leaves open the possibility that n is 22 or less could also work.

Generalizations

Arbitrary number of days

Given a year with d days, the **generalized birthday problem** asks for the minimal number $n(d)$ such that, in a set of n randomly chosen people, the probability of a birthday coincidence is at least 50%. In other words, $n(d)$ is the minimal integer n such that

Comparison of the birthday problem (1) and birthday attack (2):

In (1), collisions are found within one set, in this case, 3 out of 276 pairings of the 24 lunar astronauts.

In (2), collisions are found between two sets, in this case, 1 out of 256 pairings of only the first bytes of SHA-256 hashes of 16 variants each of benign and harmful contracts.

The classical birthday problem thus corresponds to determining $n(365)$. The first 99 values of $n(d)$ are given here (sequence A033810 in the OEIS):

d	1–2	3–5	6–9	10–16	17–23	24–32	33–42	43–54	55–68	69–82	83–99
$n(d)$	2	3	4	5	6	7	8	9	10	11	12

A similar calculation shows that $n(d) = 23$ when d is in the range 341–372.

A number of bounds and formulas for $n(d)$ have been published.[10] For any $d \geq 1$, the number $n(d)$ satisfies[11]

These bounds are optimal in the sense that the sequence $n(d) - \sqrt{2d \ln 2}$ gets arbitrarily close to

while it has

as its maximum, taken for $d = 43$.

The bounds are sufficiently tight to give the exact value of $n(d)$ in most of the cases. For example, for $d = 365$ these bounds imply that $22.7633 < n(365) < 23.7736$ and 23 is the only integer in that range. In general, it follows from these bounds that $n(d)$ always equals either

where $\lceil \cdot \rceil$ denotes the ceiling function. The formula

holds for 73% of all integers d.[12] The formula

holds for almost all d, i.e., for a set of integers d with asymptotic density 1.[12]

The formula

holds for all $d \leq 10^{18}$, but it is conjectured that there are infinitely many counterexamples to this formula.[13]

The formula

holds for all $d \leq 10^{18}$, and it is conjectured that this formula holds for all d.[13]

More than two people sharing a birthday

It is possible to extend the problem to ask how many people in a group are necessary for there to be a greater than 50% probability that at least 3, 4, 5, etc. of the group share the same birthday.

The first few values are as follows:

- >50% probability of 3 people sharing a birthday - 88 people
- >50% probability of 4 people sharing a birthday - 187 people (sequence A014088 in the OEIS).[14]

Everyone shares a birthday

The strong birthday problem asks for the number of people that need to be gathered together before there is a 50% chance that *everyone* in the gathering shares their birthday with at least one other person. For d=365 days the answer is 3,064 people.[15][16]

The number of people needed for arbitrary number of days is given by (sequence A380129 in the OEIS)

Probability of a shared birthday (collision)

The birthday problem can be generalized as follows:

> Given n random integers drawn from a discrete uniform distribution with range $[1,d]$, what is the probability $p(n; d)$ that at least two numbers are the same? ($d = 365$ gives the usual birthday problem.)[17]

The generic results can be derived using the same arguments given above.

Conversely, if $n(p; d)$ denotes the number of random integers drawn from $[1,d]$ to obtain a probability p that at least two numbers are the same, then

The birthday problem in this more generic sense applies to hash functions: the expected number of N-bit hashes that can be generated before getting a collision is not 2^N, but rather only $2^{N/2}$. This is exploited by birthday attacks on cryptographic hash functions and is the reason why a small number of collisions in a hash table are, for all practical purposes, inevitable.

The theory behind the birthday problem was used by Zoe Schnabel[18] under the name of capture-recapture statistics to estimate the size of fish population in lakes. The birthday problem and its generalizations are also useful tools for modelling coincidences.[19]

Probability of a unique collision

The classic birthday problem allows for more than two people to share a particular birthday or for there to be matches on multiple days. The probability that among n people there is exactly one pair of individuals with a matching birthday given d possible days is[19]

Unlike the standard birthday problem, as n increases the probability reaches a maximum value before decreasing. For example, for $d = 365$, the probability of a unique match has a maximum value of 0.3864 occurring when $n = 28$.

Generalization to multiple types of people

The basic problem considers all trials to be of one "type". The birthday problem has been generalized to consider an arbitrary number of types.[20] In the simplest extension there are two types of people, say m men and n women, and the problem becomes characterizing the probability of a shared birthday between at least one man and one woman. (Shared birthdays between two men or two women do not count.) The probability of no shared birthdays here is

Plot of the probability of at least one shared birthday between at least one man and one woman

where $d = 365$ and S_2 are Stirling numbers of the second kind. Consequently, the desired probability is $1 - p_0$.

This variation of the birthday problem is interesting because there is not a unique solution for the total number of people $m + n$. For example, the usual 50% probability value is realized for both a 32-member group of 16 men and 16 women and a 49-member group of 43 women and 6 men.

Other birthday problems

First match

A related question is, as people enter a room one at a time, which one is most likely to be the first to have the same birthday as someone already in the room? That is, for what n is $p(n) - p(n-1)$ maximum? The answer is 20—if there is a prize for first match, the best position in line is 20th.

Same birthday as you

In the birthday problem, neither of the two people is chosen in advance. By contrast, the probability $q(n)$ that *at least one other person* in a room of n other people has the same birthday as a *particular* person (for example, you) is given by

and for general d by

Comparing $p(n)$ = probability of a birthday match with $q(n)$ = probability of matching *your* birthday

In the standard case of $d = 365$, substituting $n = 23$ gives about 6.1%, which is less than 1 chance in 16. For a greater than 50% chance that *at least* one other person in a roomful of n people has the same birthday as *you*, n would need to be at least 253. This number is significantly higher than $\frac{365}{2} = 182.5$: the reason is that it is likely that there are some birthday matches among the other people in the room.

Number of people with a shared birthday

For any one person in a group of n people the probability that he or she shares his birthday with someone else is , as explained above. The expected number of people with a shared (non-unique) birthday can now be calculated easily by multiplying that probability by the number of people (n), so it is:

(This multiplication can be done this way because of the linearity of the expected value of indicator variables). This implies that the expected number of people with a non-shared (unique) birthday is:

Similar formulas can be derived for the expected number of people who share with three, four, etc. other people.

Number of people until every birthday is achieved

The expected number of people needed until every birthday is achieved is called the Coupon collector's problem. It can be calculated by nH_n, where H_n is the nth harmonic number. For 365 possible dates (the birthday problem), the answer is 2365.

Near matches

Another generalization is to ask for the probability of finding at least one pair in a group of n people with birthdays within k calendar days of each other, if there are d equally likely birthdays.[21]

The number of people required so that the probability that some pair will have a birthday separated by k days or fewer will be higher than 50% is given in the following table:

k	n for $d = 365$
0	23
1	14
2	11
3	9
4	8
5	8
6	7
7	7

Thus in a group of just seven random people, it is more likely than not that two of them will have a birthday within a week of each other.[21]

Number of days with a certain number of birthdays

Number of days with at least one birthday

The expected number of different birthdays, i.e. the number of days that are at least one person's birthday, is:

This follows from the expected number of days that are no one's birthday:

which follows from the probability that a particular day is no one's birthday, $\left(\frac{d-1}{d}\right)^n$, easily summed because of the linearity of the expected value.

For instance, with $d = 365$, you should expect about 21 different birthdays when there are 22 people, or 46 different birthdays when there are 50 people. When there are 1000 people, there will be around 341 different birthdays (24 unclaimed birthdays).

Number of days with at least two birthdays

The above can be generalized from the distribution of the number of people with their birthday on any particular day, which is a Binomial distribution with probability $\frac{1}{d}$. Multiplying the relevant probability by d will then give the expected number of days. For example, the expected number of days which are shared; i.e. which are at least two (i.e. not zero and not one) people's birthday is:

Number of people who repeat a birthday

The probability that the kth integer randomly chosen from $[1,d]$ will repeat at least one previous choice equals $q(k-1;\ d)$ above. The expected total number of times a selection will repeat a previous selection as n such integers are chosen equals[22]

This can be seen to equal the number of people minus the expected number of different birthdays.

The distribution of the random variable reporting the number k of integers to be chosen in order to get exactly c repeats (for a constant c) converges to a chi-distributed random variable as .[23]

Average number of people to get at least one shared birthday

In an alternative formulation of the birthday problem, one asks the *average* number of people required to find a pair with the same birthday. If we consider the probability function Pr[n people have at least one shared birthday], this *average* is determining the mean of the distribution, as opposed to the customary formulation, which asks for the median. The problem is relevant to several hashing algorithms analyzed by Donald Knuth in his book *The Art of Computer Programming*. It may be shown[24][25] that if one samples uniformly, with replacement, from a population of size M, the number of trials required for the first repeated sampling of *some* individual has expected value $\overline{n} = 1 + Q(M)$, where

The function

has been studied by Srinivasa Ramanujan and has asymptotic expansion:

With $M = 365$ days in a year, the average number of people required to find a pair with the same birthday is $\overline{n} = 1 + Q(M) \approx 24.61659$, somewhat more than 23, the number required for a 50% chance. In the best case, two people will suffice; at worst, the maximum possible number of $M + 1 = 366$ people is needed; but on average, only 25 people are required

An analysis using indicator random variables can provide a simpler but approximate analysis of this problem.[26] For each pair (i, j) for k people in a room, we define the indicator random variable X_{ij}, for , by

Let X be a random variable counting the pairs of individuals with the same birthday.

For $n = 365$, if $k = 28$, the expected number of pairs of individuals with the same birthday is $\frac{28 \times 27}{2 \times 365} \approx 1.0356$. Therefore, we can expect at least one matching pair with at least 28 people.

In the 2014 FIFA World Cup, each of the 32 squads had 23 players. An analysis of the official squad lists suggested that 16 squads had pairs of players sharing birthdays, and of these 5 squads had two pairs: Argentina, France, Iran, South Korea and Switzerland each had two pairs, and Australia, Bosnia and Herzegovina, Brazil, Cameroon, Colombia, Honduras, Netherlands, Nigeria, Russia, Spain and USA each with one pair.[27]

Voracek, Tran and Formann showed that the majority of people markedly overestimate the number of people that is necessary to achieve a given probability of people having the same birthday, and markedly underestimate the probability of people having the same birthday when a specific sample size is given.[28] Further results showed that psychology students and women did better on the task than casino visitors/personnel or men, but were less confident about their estimates.

Partition problem

A related problem is the partition problem, a variant of the knapsack problem from operations research. Some weights are put on a balance scale; each weight is an integer number of grams randomly chosen between one gram and one million grams (one tonne). The question is whether one can usually (that is, with probability close to 1) transfer the weights between the left and right arms to balance the scale. (In case the sum of all the weights is an odd number of grams, a discrepancy of one gram is allowed.) If there are only two or three weights, the answer is very clearly no; although there are some combinations which work, the majority of randomly selected combinations of three weights do not. If there are very many weights, the answer is clearly yes. The question is, how many are just sufficient? That is, what is the number of weights such that it is equally likely for it to be possible to balance them as it is to be impossible?

Often, people's intuition is that the answer is above 100 000. Most people's intuition is that it is in the thousands or tens of thousands, while others feel it should at least be in the hundreds. The correct answer is 23.

The reason is that the correct comparison is to the number of partitions of the weights into left and right. There are 2^{N-1} different partitions for N weights, and the left sum minus the right sum can be thought of as a new random quantity for each partition. The distribution of the sum of weights is approximately Gaussian, with a peak at $500\,000N$ and width $1\,000\,000\sqrt{N}$, so that when 2^{N-1} is approximately equal to $1\,000\,000\sqrt{N}$ the transition occurs. 2^{23-1} is about 4 million, while the width of the distribution is only 5 million.[29]

In fiction

Arthur C. Clarke's 1961 novel *A Fall of Moondust* contains a section where the main characters, trapped underground for an indefinite amount of time, are celebrating a birthday and find themselves discussing the validity of the birthday problem. As stated by a physicist passenger: "If you have a group of more than twenty-four people, the odds are better than even that two of them have the same birthday." Eventually, out of 22 present, it is revealed that two characters share the same birthday, May 23.

Notes

1. In his autobiography, Halmos criticized the form in which the birthday paradox is often presented, in terms of numerical computation. He believed that it should be used as an example in the use of more abstract mathematical concepts. He wrote:

   > The reasoning is based on important tools that all students of mathematics should have ready access to. The birthday problem used to be a splendid illustration of the advantages of pure thought over mechanical manipulation; the inequalities can be obtained in a minute or two, whereas the multiplications would take much longer, and be much more subject to error, whether

the instrument is a pencil or an old-fashioned desk computer. What calculators do not yield is understanding, or mathematical facility, or a solid basis for more advanced, generalized theories.

References

1. David Singmaster, *Sources in Recreational Mathematics: An Annotated Bibliography*, Eighth Preliminary Edition, 2004, section 8.B (https://www.puzzlemuseum.com/singma/singma6/SOURCES/singma-sources-edn8-2004-03-19.htm#_Toc69534221)
2. H.S.M. Coxeter, "Mathematical Recreations and Essays, 11th edition", 1940, p 45, as reported in I. J. Good, *Probability and the weighing of evidence*, 1950, p. 38 (https://archive.org/details/probabilityweigh0000good/page/38/mode/2up?q=same%20birthday)
3. Richard Von Mises, "Über Aufteilungs- und Besetzungswahrscheinlichkeiten", *Revue de la faculté des sciences de l'Université d'Istanbul* **4**:145-163, 1939, reprinted in Frank, P.; Goldstein, S.; Kac, M.; Prager, W.; Szegö, G.; Birkhoff, G., eds. (1964). *Selected Papers of Richard von Mises*. Vol. 2. Providence, Rhode Island: Amer. Math. Soc. pp. 313–334.
4. see Birthday#Distribution through the year
5. (Bloom 1973)
6. Steele, J. Michael (2004). *The Cauchy-Schwarz Master Class* (https://archive.org/details/cauchyschwarzmas00stee_431). Cambridge: Cambridge University Press. pp. 206 (https://archive.org/details/cauchyschwarzmas00stee_431/page/n217), 277. ISBN 9780521546775.
7. Mario Cortina Borja; John Haigh (September 2007). "The Birthday Problem" (https://doi.org/10.1111%2Fj.1740-9713.2007.00246.x). *Significance*. **4** (3). Royal Statistical Society: 124–127. doi:10.1111/j.1740-9713.2007.00246.x (https://doi.org/10.1111%2Fj.1740-9713.2007.00246.x).
8. Mathis, Frank H. (June 1991). "A Generalized Birthday Problem" (http://http.cs.berkeley.edu/~daw/papers/genbday-crypto02.ps). *SIAM Review*. **33** (2): 265–270. doi:10.1137/1033051 (https://doi.org/10.1137%2F1033051). ISSN 0036-1445 (https://search.worldcat.org/issn/0036-1445). JSTOR 2031144 (https://www.jstor.org/stable/2031144). OCLC 37699182 (https://search.worldcat.org/oclc/37699182).
9. Jim Gray, Catharine van Ingen. Empirical Measurements of Disk Failure Rates and Error Rates (https://arxiv.org/abs/cs/0701166)
10. D. Brink, A (probably) exact solution to the Birthday Problem, Ramanujan Journal, 2012, [1] (https://link.springer.com/article/10.1007/s11139-011-9343-9).
11. Brink 2012, Theorem 2
12. Brink 2012, Theorem 3
13. Brink 2012, Table 3, Conjecture 1
14. "Minimal number of people to give a 50% probability of having at least n coincident birthdays in one year" (https://oeis.org/A014088). *The On-line Encyclopedia of Integer Sequences*. OEIS. Retrieved 17 February 2020.
15. DasGupta, Anirban. "The matching, birthday and the strong birthday problem: a contemporary review." Journal of Statistical Planning and Inference 130.1-2 (2005): 377-389.
16. Mario Cortina Borja, The Strong Birthday Problem, Significance, Volume 10, Issue 6, December 2013, Pages 18–20, https://doi.org/10.1111/j.1740-9713.2013.00705.x
17. Suzuki, K.; Tonien, D.; et al. (2006). "Birthday Paradox for Multi-collisions". In Rhee M.S., Lee B. (ed.). *Lecture Notes in Computer Science, vol 4296*. Berlin: Springer. doi:10.1007/11927587_5 (https://doi.org/10.1007%2F11927587_5). Information Security and Cryptology – ICISC 2006.
18. Z. E. Schnabel (1938) *The Estimation of the Total Fish Population of a Lake*, American Mathematical Monthly **45**, 348–352.
19. M. Pollanen (2024) *A Double Birthday Paradox in the Study of Coincidences*, Mathematics **23**(24), 3882. https://doi.org/10.3390/math12243882
20. M. C. Wendl (2003) *Collision Probability Between Sets of Random Variables (https://dx.doi.org/10.1016/S0167-7152(03)00168-8)*, Statistics and Probability Letters **64**(3), 249–254.
21. M. Abramson and W. O. J. Moser (1970) *More Birthday Surprises*, American Mathematical Monthly **77**, 856–858
22. Might, Matt. "Collision hash collisions with the birthday paradox" (http://matt.might.net/articles/counting-hash-collisions/). *Matt Might's blog*. Retrieved 17 July 2015.
23. Corollary 5 in Arratia, Richard; Garibaldi, Skip; Kilian, Joe (2016). "Asymptotic Distribution for the Birthday Problem with Multiple Coincidences, via an Embedding of the Collision Process". *Random Structures & Algorithms*. **48** (3): 480–502.
24. Knuth, D. E. (1973). *The Art of Computer Programming*. Vol. 3, Sorting and Searching. Reading, Massachusetts: Addison-Wesley. ISBN 978-0-201-03803-3.
25. Flajolet, P.; Grabner, P. J.; Kirschenhofer, P.; Prodinger, H. (1995). "On Ramanujan's Q-Function" (https://doi.org/10.1016%2F0377-0427%2893%29E0258-N). *Journal of Computational and Applied Mathematics*. **58**: 103–116. doi:10.1016/0377-0427(93)E0258-N (https://doi.org/10.1016%2F0377-0427%2893%29E0258-N).
26. Cormen; et al. *Introduction to Algorithms*.
27. Fletcher, James (16 June 2014). "The birthday paradox at the World Cup" (https://www.bbc.co.uk/news/magazine-27835311). *bbc.com*. BBC. Retrieved 27 August 2015.
28. Voracek, M.; Tran, U. S.; Formann, A. K. (2008). "Birthday and birthmate problems: Misconceptions of probability among psychology undergraduates and casino visitors and personnel". *Perceptual and Motor Skills*. **106** (1): 91–103. doi:10.2466/pms.106.1.91-103 (https://doi.org/10.2466%2Fpms.106.1.91-103). PMID 18459359 (https://pubmed.ncbi.nlm.nih.gov/18459359). S2CID 22046399 (https://api.semanticscholar.org/CorpusID:22046399).
29. Borgs, C.; Chayes, J.; Pittel, B. (2001). "Phase Transition and Finite Size Scaling in the Integer Partition Problem". *Random Structures and Algorithms*. **19** (3–4): 247–288. doi:10.1002/rsa.10004 (https://doi.org/10.1002%2Frsa.10004). S2CID 6819493 (https://api.semanticscholar.org/CorpusID:6819493).

Bibliography

- Abramson, M.; Moser, W. O. J. (1970). "More Birthday Surprises". *American Mathematical Monthly*. **77** (8): 856–858. doi:10.2307/2317022 (https://doi.org/10.2307%2F2317022). JSTOR 2317022 (https://www.jstor.org/stable/2317022).
- Bloom, D. (1973). "A Birthday Problem". *American Mathematical Monthly*. **80** (10): 1141–1142. doi:10.2307/2318556 (https://doi.org/10.2307%2F2318556). JSTOR 2318556 (https://www.jstor.org/stable/2318556).
- Kemeny, John G.; Snell, J. Laurie; Thompson, Gerald (1957). *Introduction to Finite Mathematics* (First ed.).
- McKinney, E. H. (1966). "Generalized Birthday Problem". *American Mathematical Monthly*. **73** (5): 385–387. doi:10.2307/2315408 (https://doi.org/10.2307%2F2315408). JSTOR 2315408 (https://www.jstor.org/stable/2315408).
- Mosteller, F. (1962). "Understanding the Birthday Problem". *The Mathematics Teacher*. **55** (5): 322–325. doi:10.5951/MT.55.5.0322 (https://doi.org/10.5951%2FMT.55.5.0322). JSTOR 27956609 (https://www.jstor.org/stable/27956609). Reprinted in Mosteller, Frederick (2006). "Understanding the Birthday Problem". *Selected Papers of Frederick Mosteller*. Springer Series in Statistics. pp. 349–353. doi:10.1007/978-0-387-44956-2_21 (https://doi.org/10.1007%2F978-0-387-44956-2_21). ISBN 978-0-387-20271-6.
- Schneps, Leila; Colmez, Coralie (2013). "Math error number 5. The case of Diana Sylvester: cold hit analysis". *Math on Trial. How Numbers Get Used and Abused in the Courtroom*. Basic Books. ISBN 978-0-465-03292-1.
- Sy M. Blinder (2013). *Guide to Essential Math: A Review for Physics, Chemistry and Engineering Students* (https://books.google.com/books?id=M7TCNAEACAAJ). Elsevier. pp. 5–6. ISBN 978-0-12-407163-6.

External links

- The Birthday Paradox accounting for leap year birthdays (http://www.efgh.com/math/birthday.htm)
- Weisstein, Eric W. "Birthday Problem" (https://mathworld.wolfram.com/BirthdayProblem.html). *MathWorld*.
- A humorous article explaining the paradox (http://www.damninteresting.com/?p=402)
- SOCR EduMaterials activities birthday experiment (http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_BirthdayExperiment) Archived (https://web.archive.org/web/20180412152856/http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_BirthdayExperiment) 2018-04-12 at the Wayback Machine
- Understanding the Birthday Problem (Better Explained) (http://betterexplained.com/articles/understanding-the-birthday-paradox/)
- Eurobirthdays 2012. A birthday problem. (http://www.matifutbol.com/en/eurobirthdays.html) A practical football example of the birthday paradox.
- Grime, James. "23: Birthday Probability" (https://web.archive.org/web/20170225140726/http://www.numberphile.com/videos/23birthday.html). *Numberphile*. Brady Haran. Archived from the original (http://www.numberphile.com/videos/23birthday.html) on 2017-02-25. Retrieved 2013-04-02.
- Computing the probabilities of the Birthday Problem at WolframAlpha (https://www.wolframalpha.com/input/?i=birthday+paradox%2C+4+people%2C+100+possible+birthdays)