How Likely Is It that Two Classmates Have the Same Birthday?

Sergio Fantini and Jeffrey C. Grossman

Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801

What is the probability that at least two students in a class of 30 were born on the same day of the year? The answer to this apparently simple question is quite surprising and goes against our common sense. A solution to this problem can be obtained[1] by first deriving the probability that everyone in the class has a different birth date. Thus, for a given birthday of the first student, the chances that the second student has a different birthday are 364/365, and the chances that the third has a different birthday than the first two are 363/365, and so on. In order to obtain the probability that all of the students have different birth dates, we simply multiply each of these individual probabilities together. The probability, $p(30)$, that in a class of 30 there are at least two students with the same birthday is then given by one minus the probability of having all different birthdays:

$$p(30) = 1 - \frac{364}{365}\frac{363}{365}\frac{362}{365}\cdots\frac{336}{365} = 70.63\% \qquad (1)$$

Such a large probability is truly surprising and much higher than our intuition dictates. In order to further investigate this seemingly anomalous result, we probe the multitude of probability components in somewhat more detail.

Let us begin by considering the general case in which the number of students is m and the number of days in a year is N [by letting both m and N be generic, the final equation will also be applicable to other problems such as finding the probability that rolling m dice ($N = 6$) gives rise to at least a two-of-a-kind]. We want to find the probability $p(m)$ by taking the ratio between the number of combinations in which at least two birthdays coincide and the total number of combinations of birthdays. For the simple case of just two students, the total number of combinations of birthdays is $(365)^2$. This is easily generalized to the case of N days in the year and m students, which gives a total of N^m. To find the number of combinations in which at least two birthdays coincide, we start by considering the number of ways in which the first student can share her birthday with at least one other student. In a class of m students this number is given by:

$$[N^{m-1} - (N-1)^{m-1}]\,N \qquad (2)$$

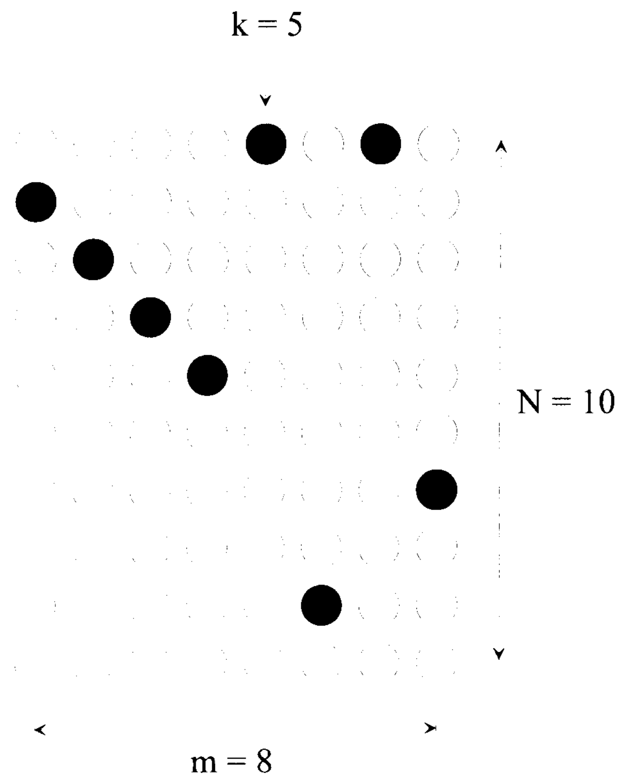

Fig. 1. Schematic representation of counting strategy used to calculate probability $p\,(m)$. Each column represents the N possible birthdays for a given student, and there are m columns (as many as the number of students). In this example $m = 8$ and $N = 10$ for simplicity. The birthday for each student is represented by a filled circle. The generic term k in Eq. (4) is obtained by counting the possible combinations where the k^{th} student has the same birthday as at least one successive student, while the preceding students do not share their birthday with any other student. Figure shows one such combination.

where N^{m-1} is the total number of combinations of the remaining $m - 1$ students, and $(N - 1)^{m-1}$ is the number of cases in which no students were born the same day of student 1. The factor N in Eq. (2) accounts for the N possibilities for the birthday of student 1.

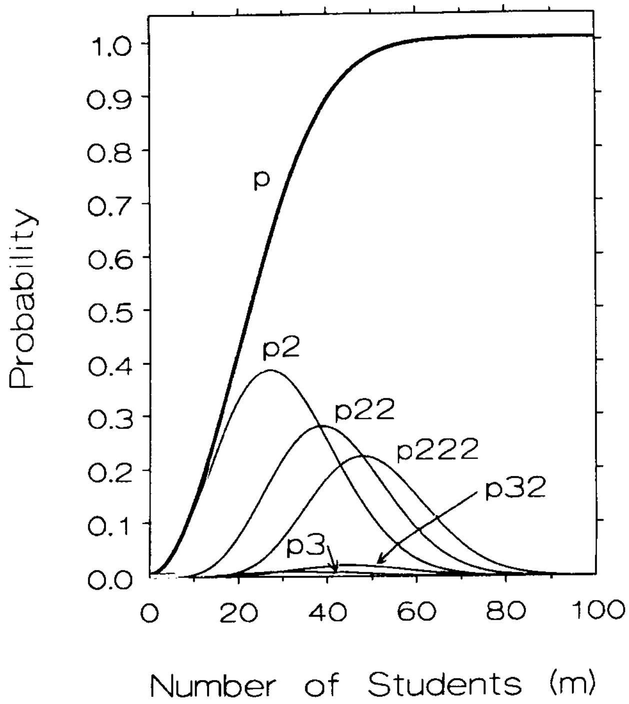

Fig. 2. Probabilities as a function of number of students in the class. p, p_2, p_3, p_{22}, p_{32}, and p_{222} correspond to the probability of having the following combinations of students with the same birthday: at least two, exactly two, exactly three, exactly two groups of two, exactly one group of three and one group of two, and exactly three groups of two, respectively.

The next step is to find the number of combinations in which student 2 was born on the same day as at least one other student (excluding the cases of being born on the same day as student 1 since those have already been counted):

$$[(N-1)^{m-2} - (N-2)^{m-2}]\, N\,(N-1) \qquad (3)$$

Here, $(N-1)^{m-2}$ is the number of combinations of the $m-2$ students (we do not consider students 1 and 2) in which no students were born the same day as student 1, and $(N-2)^{m-2}$ is the number of ways in which no students were born the same day as either student 1 or student 2. The factor $N(N-1)$ in Eq. (3) accounts for the N possibilities for the birthdays of student 2, and the $N-1$ possibilities for the birthday of student 1 (remember that the birthdays of students 1 and 2 must be different).

By proceeding in this way, we find that the general number of combinations in which student k was born the same day as at least one other student, *and* students 1, 2, 3,..., $(k-1)$ do not have their birthday in common with any other classmate is:

$$[(N-k+1)^{m-k} - (N-k)^{m-k}]\, N\,(N-1)(N-2)\ldots(N-k+1) \qquad (4)$$

where k must be less than or equal to the minimum between $m-1$ and N (if $k = m$ there are 0 combinations in which the last student shares the birthday with a successive student, while if $k = N+1$ there are 0 combinations in which the k^{th} student does not share a birthday with the preceding student). This counting scheme is illustrated in Fig. 1 for the particular case $m = 8$, $N = 10$, $k = 5$. As already mentioned, the probability $p(m)$ is given by the ratio between the number of combinations where at least two birthdays coincide [summation over all students of Eq. (4)], and the total number of combinations of birthdays (N^m). Hence, the requested probability $p(m)$ is:

$$p(m) = \frac{1}{N^m} \sum_{k=1}^{min(m-1,\, N)} \frac{N!}{(N-k)!} [(N-k+1)^{m-k} - (N-k)^{m-k}] \qquad (5)$$

where we have written $N(N-1)(N-2)\ldots(N-k+1) = N!/(N-k)!$.

The probability given by Eq. (5) is shown by the red line in Fig. 2 as a function of the number of students m. First of all, we see that our original answer using the simpler derivation is recovered, namely that the probability that at least two students in a class of 30 ($m = 30$) were born the same day of the year ($N = 365$) is 70.63%. Perhaps even more surprising than this result is the fact that a class of 60 students has a probability of 99.41%, and a class of 80 students has a probability of 99.99%. Of course, only a class of 366 students has a probability of exactly 100%. In practice, however, in a group of 80 randomly selected students at least two of them will almost always have the same birthday.

But the surprises are not over. Using this more complicated and detailed analysis has enabled us to take an even closer look at the problem. The probability $p(m)$ of Eq. (5) tells us the odds

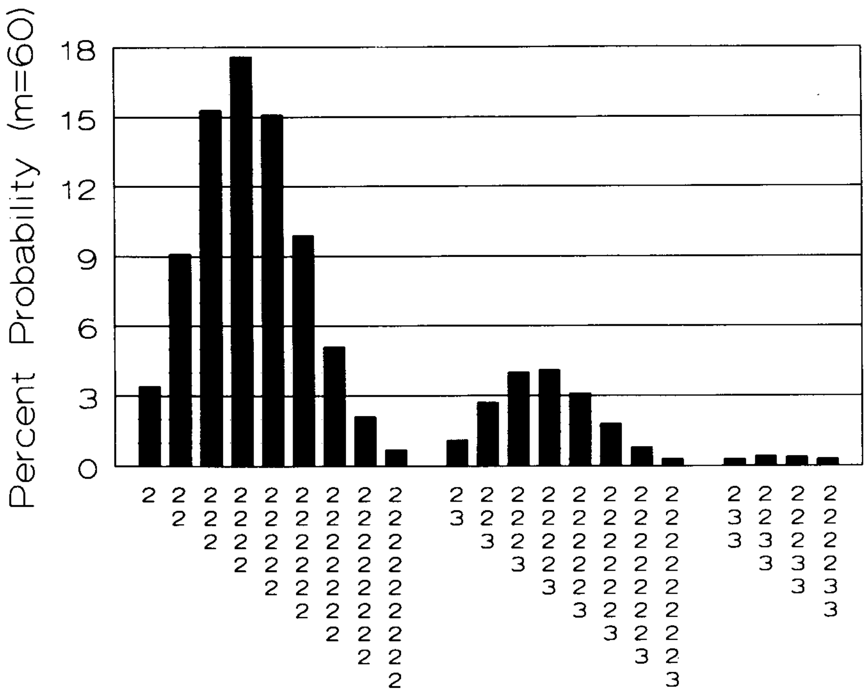

Fig. 3. Percent probability of having different combinations of groups of students with the same birthday in a class of 60, calculated with a Monte Carlo sampling of 10^6 classes. Labels on *x*-axis refer to different combinations of groups. For instance, 2223 means that there are four groups of students, namely three pairs and one group of three students, who share the same birthday (so that there are four distinct birthdays involved). The probability of having at least two coincident birthdays is $p(60) = 99.4\%$. The sum of the contributions shown in the figure is 97.6%. The missing 1.8% is distributed among many more, smaller contributions.

that *at least* two students have the same birthday in a class of m students. It is interesting to analyze some of the various components of $p(m)$ by looking at the probabilities of having exactly two students (p_2), exactly three students (p_3), exactly two pairs of students (p_{22}), exactly two groups, one of 3 and one of 2 students (p_{32}), and exactly three pairs of students (p_{222}) who share the same birthday. These probabilities can be found with similar procedures as the one used to find p. The results are the following:

$$p_2(m) = \frac{1}{N^m}\frac{N!}{(N-m+1)!}\frac{1}{2}m(m-1) \tag{6}$$

$$p_3(m) = \frac{1}{N^m}\frac{N!}{(N-m+2)!}\frac{1}{6}m(m-1)(m-2) \tag{7}$$

$$p_{22}(m) = \frac{1}{N^m}\frac{N!}{(N-m+2)!}\frac{1}{8}m(m-1)(m-2)(m-3) \tag{8}$$

$$p_{32}(m) = \frac{1}{N^m}\frac{N!}{(N-m+3)!}\frac{1}{12}m(m-1)(m-2)(m-3)(m-4) \tag{9}$$

$$p_{222}(m) = \frac{1}{N^m}\frac{N!}{(N-m+3)!}\frac{1}{48}m(m-1)(m-2)(m-3)(m-4)(m-5) \tag{10}$$

where the factorial of a negative integer is conventionally set to infinity.[2] These probabilities are shown in Fig. 2 as a function of the number of students m, together with the probability $p(m)$. One surprising result is that not only does a group of 80 students have a 99.99% probability of having at least two students with the same birthday, but the chances of having just two, or two pairs, or three pairs is negligibly small. In this case ($m = 80$), it is almost certain that there is more than one group of students with the same birthday. So where exactly are the contributions to $p(m)$ coming from?

To answer such a question, we have investigated the probability contributions to $p(m)$ for the case of 60 students. The value of $p(60)$ is 99.41%, but the sum of the other probabilities shown in Fig. 2 is only 29.13%. With the Monte Carlo approach, based on a sampling of 10^6 classes, we have been able to quantify all the contributions to $p(60)$ beyond those reported in Eqs. (6) through (10). In Fig. 3 we report the main contributions, each shown as a probability percentage. Thus, we see that for 60 students the highest probability is p_{2222}, although p_{22222}, p_{222222}, p_{322}, and p_{3222} all have significant contributions. The cases that include one group of four have a very minor contribution. In general, for an arbitrary class size (even up to several hundred students), most of the contributions involve multiple pairs, and almost all of the pieces include *at least* one pair.

We conclude by noting that in a group of 80 students, the most likely situation (probability of 10.3%) is that there are seven pairs of students with the same birthday!

References

1. G. Gamow, *One Two Three...Infinity* (Dover, New York, 1991), pp. 213–215.
2. G. Arfken, *Mathematical Methods for Physicists* (Academic Press, San Diego, CA, 1985), p. 544.