Now consider an arbitrary scheme with volume $V$. For any $\epsilon>0$ we can find an integer $n$ so that the boxes from the first $n$ stages of this scheme have volume at least $V-\epsilon$. Since these boxes belong to an $n$-stage problem, their volume is at most $\bar{V}_n$ and therefore less than $\bar{V}$. It follows that $V<\bar{V}+\epsilon$ and, since $\epsilon$ is arbitrary, $V\leqslant\bar{V}$. This shows that $\bar{V}$ is indeed the maximum volume in the infinite-stage problem.

Finally, if a scheme produces the maximum volume $\bar{V}$, then all of the subschemes belonging to any stage must produce $\bar{V}$ times their scaling factor or they could be replaced with a definite improvement. Without loss of generality we consider the $\mu$ of the first stage. It must satisfy

$$\bar{V}=\mu(1-2\mu)^2+4\mu^3\bar{V}$$

or

$$4(\bar{V}+1)\mu^3-4\mu^2+\mu-\bar{V}=0,$$

and this equation suffices to prove that $\mu=\bar{\lambda}$. This follows because $\bar{\lambda}$ is a double root and therefore the remaining root must be

$$\frac{\bar{V}}{4(\bar{V}+1)\bar{\lambda}^2}=.5827\ldots>\frac{1}{2}.$$

This completes the proof that the only scheme with volume $\bar{V}$ is the one in which $\mu_i=\bar{\lambda}$ for all $i$.

# The Uniformity Assumption in the Birthday Problem

GEOFFREY C. BERRESFORD
*Long Island University*
*Greenvale, NY 11548*

The birthday problem, to find the probability that in a group of $n$ people some two will share a common birthday, has occurred frequently in the literature since having been proposed in 1939 by von Mises. It is easily solved under the assumptions that each person's birthday is determined independently, and that the 365 possible birthdays (ignoring leap years) are equally likely. Under these independence and uniformity assumptions it is easy to show that the probability of a shared birthday reaches $\frac{1}{2}$ as soon as the size of the group reaches 23.

The reason for the uniformity assumption is interesting. Depending on the population, it may or may not be a reasonable approximation to reality, but in any case it is enormously convenient. To see this, let us consider the problem without assuming uniformity. To take full account of the 365 probabilities of being born on different days of the year, we let $p_i$ be the probability of being born on day $i$, $i=1,\ldots,365$, and obtain the (complementary) probability of $n$ independently chosen people all having different birthdays as

$$P(n)=n!\sum_{i_1<\cdots<i_n}p_{i_1}\cdots p_{i_n},\tag{1}$$

the sum being over all $n$-subsets of $\{1,2,\ldots,365\}$ such that $i_1<i_2<\cdots<i_n$. The difficulty is that the sum has $\binom{365}{n}$ terms, and for group size $n=23$ this is $\binom{365}{23}\approx10^{36}$ terms, which even the fastest computer would need $10^{20}$ centuries to calculate.

Nevertheless, two observations can be made, one theoretical, one empirical. *In a group of n people, the probability of a shared birthday is least for the uniform distribution.* Therefore, regardless of the actual distribution of birthdays, a group size of 23 is sufficient to make a shared birthday more probable than not. There are proofs of this in the literature, but the following version of Munford's 1977 proof [3] is perhaps the simplest. We will show that the (complementary) probability $P(n)$ of $n$ independently chosen people having different birthdays is greatest for the uniform distribution. We assume, of course, that $n \geqslant 2$ and that at least $n$ of the $p_i$'s are nonzero, so that $n$ different birthdays are possible. We will show how $P(n)$ changes if we alter the values of two unequal probabilities, say $p_1$ and $p_2$, by replacing them with their common mean, $\frac{1}{2}(p_1+p_2)$. Let us partition the sum (1) into three parts: those terms containing both $p_1$ and $p_2$, those terms containing just one of $p_1$ and $p_2$, and those terms containing neither $p_1$ nor $p_2$. Let us also factor out $p_1$ and $p_2$ whenever they occur. We may then express $P(n)$ as

$$P(n)=n!\left(p_1 p_2 \sum_{2<i_1<\cdots<i_{n-2}} p_{i_1}\cdots p_{i_{n-2}}+(p_1+p_2)\sum_{2<i_1<\cdots<i_{n-1}} p_{i_1}\cdots p_{i_{n-1}}\right.$$
$$\left.+\sum_{2<i_1<\cdots<i_n} p_{i_1}\cdots p_{i_n}\right). \tag{2}$$

If we now replace both $p_1$ and $p_2$ by their common mean, $\frac{1}{2}(p_1+p_2)$, thus leaving their sum unchanged, only the first term in (2) changes, in which we must replace $p_1 p_2$ by $(\frac{1}{2}(p_1+p_2))^2$. But since $p_1 p_2<(\frac{1}{2}(p_1+p_2))^2$, (taking the square root of each side, this is merely the statement that the geometric mean of two unequal numbers is less than their arithmetic mean) this replacement will only increase the value of $P(n)$. Thus since the probability $P(n)$ of different birthdays can be increased by this operation whenever two $p_i$'s are unequal, it must be greatest when all the $p_i$'s are equal, that is, for the uniform distribution.

The second observation involves comparing the uniformity assumption with actual data. FIGURE 1 is a graph of the empirical probabilities of a birth occurring on any day of the year 1977 for the 239,762 live births in New York State (source: New York State Health Department). I leave it to the reader to surmise reasons for the obvious weekly cyclical component. The empirical probabilities vary from a low of .002135 (on Sunday, December 11th) to a high of .003478 (on Wednesday, July 6th), a variation of almost 27% from the mean of 1/365. Thus for a population born in a given year (the type of population from which most school classes are drawn) the assumption of uniformity is not valid. However, because a given birthday will fall on different days of the week in different years (since 365 is relatively prime to 7) in a population of mixed ages the weekly cycle will be averaged out. For such a population uniformity will be a reasonable assumption, as is shown by the graph in FIGURE 2, in which the data is as in FIGURE 1 except that each daily probability has been averaged with the six following it to remove the weekly cycle. The variation here is only about 10% above and below the mean of 1/365.

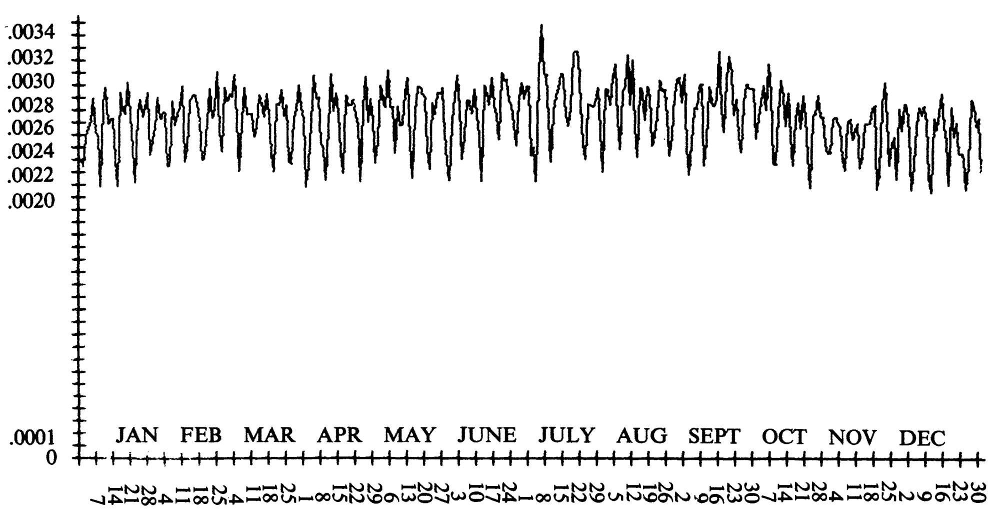


FIGURE 1.

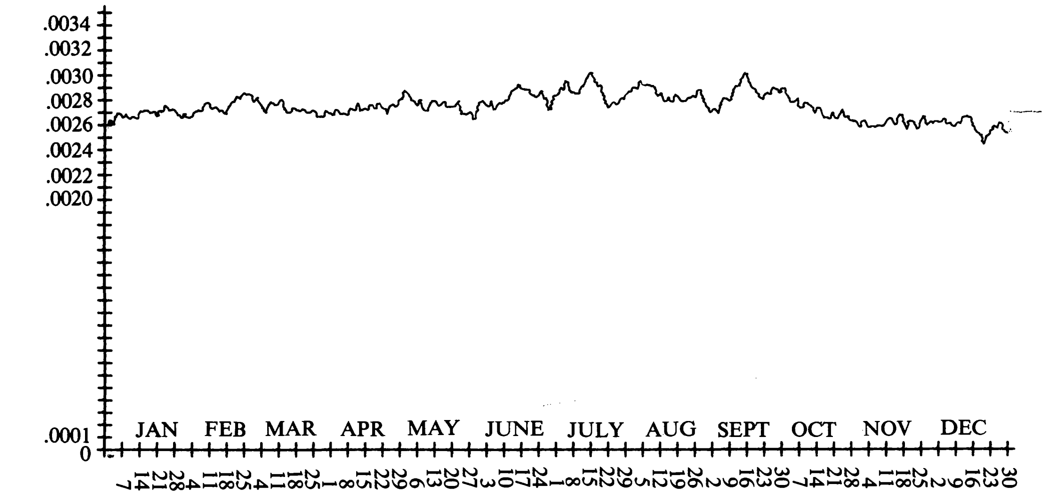


FIGURE 2.

Although the birthday problem with 365 different probabilities is hopelessly intractable, it becomes manageable if one allows a smaller number, $m$, of different probabilities, $p_1,\ldots,p_m$, and rounds each of the actual empirical probabilities to the nearest of these. Then if $d_i$ is the number of days being assigned probability $p_i$, subject to the obvious consistency relations $\sum d_i = 365$, $\sum d_i p_i = 1$, the probability of $n$ people all having different birthdays is

$$n! \sum_{n_1+\cdots+n_m=n} \prod_{i=1}^{m} \binom{d_i}{n_i} p_i^{n_i},$$

where the sum is over the $\binom{m+n-1}{n}$ different $m$-tuples $(n_1,\ldots,n_m)$ of nonnegative integers satisfying $n_1+\cdots+n_m=n$ (see Feller [2], p. 38).

| Size ($n$) | Probability of a Shared Birthday | |
|---|---|---|
| of Group | Uniform Case | Non-Uniform Case |
| 12 | .1670 | .1683 |
| 15 | .2529 | .2537 |
| 18 | .3469 | .3491 |
| 21 | .4437 | .4463 |
| 22 | .4757 | .4783 |
| 23 | .5073 | .5101 |

TABLE 1.

A computer calculation (performed on a Univac 1100/82 with 18 digit precision) using $m = 10$ different probabilities based on the empirical probabilities from the 1977 New York State data graphed in FIGURE 1 shows that the probability of a shared birthday is surprisingly robust: a group size of 23 is still required to raise the probability above one-half (see TABLE 1).

I would like to thank the C. W. Post Computer Center and the C. W. Post Research Committee for their assistance.

**References**

[1] D. M. Bloom, A birthday problem, Amer. Math. Monthly, 80 (1973) 1141–1142.
[2] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., John Wiley, New York, 1968.
[3] A. G. Munford, A note on the uniformity assumption in the birthday problem, Amer. Statist., 31, no. 3 (1977) 119.