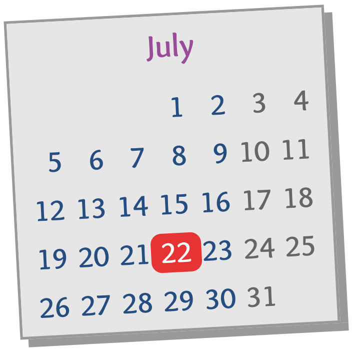
Calculating Birthday Probabilities

Paul Kotschy

19 July 2009

Compiled on February 19, 2025



Confronted by a randomly selected group of people, one would ask, "What is the probability that at least one other person in the group shares my birthday?" While this may be easy to ask, answering it is less easy. Let's begin…

For a moment, speaking more generally than birthdays, suppose the *state* of a system is completely specified by a set of n *objects*, and where each object is randomly chosen from a randomly constituted large basket of objects. Suppose too that in the basket there are only m different *types* of objects. Obviously then, each of the chosen objects can be considered to be labelled with one of the m types. In the case of birthdays then, it is easy to identify each person in the group as an object chosen from the large basket, and to identify each available birthday as an object type. That is, $m = 365$.

Define S_{mn} as the set of all possible *states*, with each state comprising n placeholders each filled with an object, and where each object is type-labelled in one of m ways. For example, if 'a', 'b' and 'c' are three such labels, then

$$
\begin{aligned}
&S_{11} = \{\{\text{'a'}\}\}, \quad S_{12} = \{\{\text{'a'}, \text{'a'}\}\}, \quad S_{13} = \{\{\text{'a'}, \text{'a'}, \text{'a'}\}\} \\
&S_{21} = \{\{\text{'a'}\}, \{\text{'b'}\}\}, \quad S_{22} = \{\{\text{'a'}, \text{'a'}\}, \{\text{'a'}, \text{'b'}\}, \{\text{'b'}, \text{'a'}\}, \{\text{'b'}, \text{'b'}\}\} \\
&S_{31} = \{\{\text{'a'}\}, \{\text{'b'}\}, \{\text{'c'}\}\}
\end{aligned}
$$

And if we relax the notational rigour, then

$$
\begin{aligned}
&S_{11} = \{a\}, \quad S_{12} = \{aa\}, \quad S_{13} = \{aaa\} \\
&S_{21} = \{a, b\}, \quad S_{22} = \{aa, ab, ba, bb\}, \quad S_{23} = \{aaa, aab, aba, abb, baa, bab, bba, bbb\} \\
&S_{31} = \{a, b, c\}, \quad S_{32} = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}
\end{aligned}
\tag{1}
$$

Let $|A|$ denote the number of elements in any set A. Then by simple inspection

$$
|S_{mn}| = m^n \tag{2}
$$

Define $T_{mn}(x)$ as the subset of S_{mn} containing all states having at least one of the n objects labelled with an 'x', and conversely, define $\overline{T}_{mn}(x)$ as the subset of S_{mn} of states not having any objects labelled with an 'x'. So, for example, by inspecting (1),

$$
\begin{aligned}
&T_{1n}(a) = S_{1n} \ \text{ for all } n \\
&T_{21}(a) = \{a\}, \quad T_{21}(b) = \{b\} \\
&T_{22}(a) = \{aa, ab, ba\}, \quad T_{22}(b) = \{ab, ba, bb\} \\
&T_{23}(a) = \{aaa, aab, aba, abb, baa, bab, bba\}, \quad T_{23}(b) = \{aab, aba, abb, baa, bab, bba, bbb\} \\
&T_{31}(a) = \{a\}, \quad T_{31}(b) = \{b\}, \quad T_{31}(c) = \{c\} \\
&T_{32}(a) = \{aa, ab, ac, ba, ca\}, \quad T_{32}(b) = \{ab, ba, bb, bc, cb\}
\end{aligned}
$$

and

$$\begin{aligned}
&\overline{T}_{1n}(a) = \{\} \text{ for all } n \\
&\overline{T}_{21}(a) = \{b\}, \quad \overline{T}_{21}(b) = \{a\} \\
&\overline{T}_{22}(a) = \{bb\}, \quad \overline{T}_{22}(b) = \{aa\} \\
&\overline{T}_{23}(a) = \{bbb\}, \quad \overline{T}_{23}(b) = \{aaa\} \\
&\overline{T}_{31}(a) = \{b, c\}, \quad \overline{T}_{31}(b) = \{a, c\}, \quad \overline{T}_{31}(c) = \{a, b\} \\
&\overline{T}_{32}(a) = \{bb, bc, cb, cc\}, \quad \overline{T}_{32}(b) = \{aa, ac, ca, cc\}, \quad \overline{T}_{32}(c) = \{aa, ab, ba, bb\}
\end{aligned}$$

Obviously

$$|T_{mn}(x)| + \left|\overline{T}_{mn}(x)\right| = |S_{mn}|$$

By definition, S_{mn} is the set of all possible ways of building a state of n filled object placeholders. So if a state was constucted randomly by drawing the objects from a large basket, say, then we may naturally ask about the chances of obtaining one or other state. In particular, we may ask for the probability of randomly obtaining a state containing at least one object labelled 'x' in any one of the n placeholders. And the answer has to be

$$p_{mn}(x) = \frac{|T_{mn}(x)|}{|S_{mn}|} \tag{3}$$

Conversely, we may ask for the probability of randomly obtaining a state containing no object labelled 'x' in any one of the placeholders. And that is

$$\overline{p}_{mn}(x) = \frac{\left|\overline{T}_{mn}(x)\right|}{|S_{mn}|} \tag{4}$$

Consider now the set $\overline{T}_{mn}(x)$. This set is nothing other than the set of all n-object states but with the number of available object types reduced by one, i.e., with type 'x' removed. Therefore

$$\left|\overline{T}_{mn}(x)\right| = \left|S_{(m-1)n}\right| \qquad \text{for any type 'x'} \tag{5}$$

Thus from Eqs. (2), (3) and (5)

$$\overline{p}_{mn}(x) = \frac{\left|S_{(m-1)n}\right|}{|S_{mn}|} = \left(\frac{m-1}{m}\right)^n$$

And since $\overline{p}_{mn}(x) + p_{mn}(x) = 1$ we have

$$\boxed{p_{mn}(x) = 1 - \left(\frac{m-1}{m}\right)^n} \tag{6}$$

Equation (6) provides the probability of finding an object of type 'x' in a random choice of n objects where each object may be type-labelled in one of m ways. The equation encapsulates the intuition that $p_{mn}(x) \to 1$ as $n \to \infty$, and that $p_{1n}(x) = 1$ for all n.

We are now equipped to answer the question, "What is the probability that somebody else in a group (at least one other person) shares my birthday?". We need simply apply (6) to a group of people and their respective birthdates. Obviously we must set $m = 365$. So if there are n people in the group, then from (6) the probability is

$$p_{365,n}(x) = 1 - \left(\frac{365-1}{365}\right)^n$$

For example, if there 10 people in the group, then the probability that at least one other person shares my birthday, which happens to be on date 'x', is

$$p_{365,10}(x) = 1 - \left(\frac{365-1}{365}\right)^{10} \approx 0.03$$

And if there are 50 people in the group, then the probability increases to

$$p_{365,50}(x) = 1 - \left(\frac{365 - 1}{365}\right)^{50} \approx 0.13$$

Equation (6) also exposes a counter-intuitive fact regarding birthday probabilities. The size of the group of people needs to be greater than $\frac{1}{2} \times 365$ to provide a 50% chance that someone will share your birthday. For if we set $m = 365$ and $n = 365/2$, then

$$p_{365,365/2} = 1 - \left(\frac{364}{365}\right)^{365/2} \approx 0.39$$

which is less than 0.5. Whereas, say,

$$p_{365,255} = 1 - \left(\frac{364}{365}\right)^{255} \approx 0.5$$