

9.9 The Birthday Problem

WORKSHEET

The question

Given that $0 < r < n$ and r is much smaller than n , show that $\frac{n-r}{n} \approx e^{-r/n}$.

There are k guests at a party. Assuming that there are exactly 365 days in the year, and that the birthday of any guest is equally likely to fall on any of these days, show that the probability that there are at least two guests with the same birthday is approximately $1 - e^{-k(k-1)/730}$.

Using the approximation $\frac{253}{365} \approx \ln 2$, find the smallest value of k such that the probability that at least two guests share the same birthday is at least $\frac{1}{2}$.

How many guests must there be at the party for the probability that at least one guest has the same birthday as the host to be at least $\frac{1}{2}$?

Hints

Try the question on your own first. If you get stuck, read the hints *one at a time*, and go back to the question after each one. Hints marked *Pause and try* are the key steps.

- ? **Ask yourself** Why is it easier to calculate the probability that *no* two guests share a birthday? Line the guests up: what is the chance that the second avoids the first's birthday? That the third avoids both?

STAGE 1 A useful approximation

- ? **Ask yourself** What is the tangent line to $y = e^{-x}$ at $x = 0$?

STAGE 2 The probability of a shared birthday

- ? **Ask yourself** Look at the exponent: $\frac{k(k-1)}{730} = \binom{k}{2} / 365$. What does that say in words?

STAGE 4 Sharing with the host

- ? **Ask yourself** For each guest, what is the probability of *not* sharing the host's birthday? Are the guests independent of each other in this respect?

Try this next

- 1 The first repeat.** (From STEP II, 2014.) A random number generator produces integers from 1 to n , independently and equally likely. Let X be the position of the first repeat. Find $P(X = r)$ for $2 \leq r \leq n + 1$, and deduce that

$$\frac{1}{n} + \left(1 - \frac{1}{n}\right)\frac{2}{n} + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\frac{3}{n} + \cdots = 1.$$

Show that $E(Y) = \sum_{k=1}^N P(Y \geq k)$ for any random variable taking values $1, \dots, N$, and hence that

$$\left(1 - \frac{1^2}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2^2}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3^2}{n}\right) + \cdots = 0.$$

- 2 Three in common.** Roughly how many people are needed for a better-than-even chance that some *three* share a birthday? (*Hint:* the expected number of triples sharing a birthday is $\binom{k}{3}/365^2$. Set this to about $\ln 2$.) The true answer is 88.
- 3 Near misses.** How many people are needed for an even chance that two have birthdays on the same day or on adjacent days?

- 4 Uneven birthdays.** Real birthdays are not uniform. Show that if the days have probabilities $p_1, \dots, p_{365}$, the chance that two people share a birthday is $\sum p_i^2$, and that this is smallest when all p_i are equal. (So non-uniformity only makes matches more likely.)

Further reading

- **William Feller**, *An Introduction to Probability Theory and Its Applications*, Vol. 1 (Wiley), Chapter II. The birthday problem and its approximations.
- **Persi Diaconis & Frederick Mosteller**, “Methods for studying coincidences”, *Journal of the American Statistical Association* 84 (1989). Why coincidences are common; very readable.
- **David Hand**, *The Improbability Principle* (Bantam). The square-root law and other reasons why unlikely things happen all the time.

The full guided solution, and what the question leads to at university, is in Section 9.9:
https://keithmcnulty.github.io/step_book_mcnulty/ch09/birthday.html