Birthday Problem 2

BIOF2014

Problem

What is the probability that at least two people in this classroom share the same birthday?

When we examined this problem previously, we assumed:

- 365 days in a year
- Every birthday is equally likely

However, these assumptions are unrealistic. Let's refine our model by using real world data.

Data

The UK publishes the average frequency of births on each day of the year from 1995 to 2024. We can download it with

```
curl -o uk-daily-births.csv \
https://www.ons.gov.uk/visualisations/nesscontent/dvc307/line_chart/data.csv
```

So, let's use this data to construct the empirical probability mass function (pmf). Then, we can re-estimate the probability that two people share the same birthday in a group of people.

Variables

Let N be the number of people in this class.
Let X be the number of people with the same birthday.
Let J be the number of days in a year.

Tasks

1. Load the data into R and visualize it.
2. Account for leap years. We can consider this by calculating the total numbers of births on each day from the average number of births on each day across the period.
3. Construct the empirical pmf that a person is born on each day within a year.
4. Implement a simulation that generates the number of people who share the same birthday based on the empirical pmf.
5. Compare $P(X \geq 2)$ determined using the simple uniform model we used previously vs. the probability that is estimated using the empirical pmf.

Reference

Blais 2014, section 2.4