Birthdays Project

Shaambhav Dave & Dahyun Hong

Table of Contents

Introduction

- Problem 3 from Problem Set 1 in Orloff and Booth presents the following scenario: Assuming that birthdays are equally likely to fall on any day of the year (ignoring leap years), consider a group of k people.

- An element of the sample space Ω will be the sequence of k birthdays (one for each person).

- Our focus is on the probability of at least m of them sharing the same birthday, using a mathematical model and simulations in R.

Background – An Overview

- Probability that among k people, at least m share the same birthday ignoring leap days and assuming all birthdays are equally likely to fall on any given day of the year

$$S = \{1, 2, 3, \ldots, n\}, n = 365$$

$$S^k = \{b = (b_1, b_2, \ldots, b_n) : b_i \in S\}$$

$$P(\{b\}) = \frac{1}{365^k}$$

$$D_k = \left\{b \in S^k : b_i \neq b_j,\ i \neq j\right\}$$

$$\#D_k = P(n, k) = \frac{n!}{(n-k)!}$$

$$M_l(k) = \left\{b \in S^k : max_{j=1,\ldots,k} \#\{i \in 1, .., k : b_i = b_j\} = l\right\}$$

Case Study: m = 2

- Calculating probability for at least 2 out of k people sharing the same birthday.
- We can start off with the base function and assume that at most one birthday is shared: $\bigcup_{l=2}^{k} M_l = S_k \setminus M_1$
- By the property of the $M_2 \subset S_k$, le $\sum_{l=m}^{k} \#M_l = n^k - \frac{n!}{(n-k)!}$ et
- Substituting into the formula for the product measure:

$$p(k;2,n) = \frac{1}{n^k}\left(n^k - \frac{n!}{(n-k)!}\right) = 1 - \frac{n!}{n^k(n-k)!}$$

Case Study: m = 3, k = 3

- To calculate probability for all three people sharing the same birthdays: $\bigcup_{l=3}^{k} M_l,\ k = 3$
- Since all 3 people share the same birthday and there are 365 possible days that the shared birthday could fall on:

$$p(3; 3,\ n) = \frac{n}{n^k} = \frac{1}{n^{k-1}} = \frac{1}{365^2} \approx 7.506 \times 10^{-6}$$

- By the Bayes Theorem, assuming each person's birthday is independent of the others:

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C) = \frac{1}{365^2}$$

Case Study: m = 3, k = 4

- Unlike the other cases, we need to split this into two parts:
  - There are exactly 3 people who share the same birthday:

$$\frac{C(4,3)\times 365\times 364}{365^{4}} = 2.994 \times 10^{-5}$$

  - There are exactly 4 people who share the same birthday:

$$\sum_{l=4}^{4} \#M_{l} = 365, \quad p(4;4,n) = \frac{1}{n^{k-1}} = \frac{1}{365^{3}} \approx 2.056 \times 10^{-8}$$

- Summing the two cases, we find: $p(4;3,n) \approx 2.996 \times 10^{-5}$

Other Considerations

- The following scenarios are presented in Orloff and Booth:
  - Event A: “someone in the group shares your birthday”

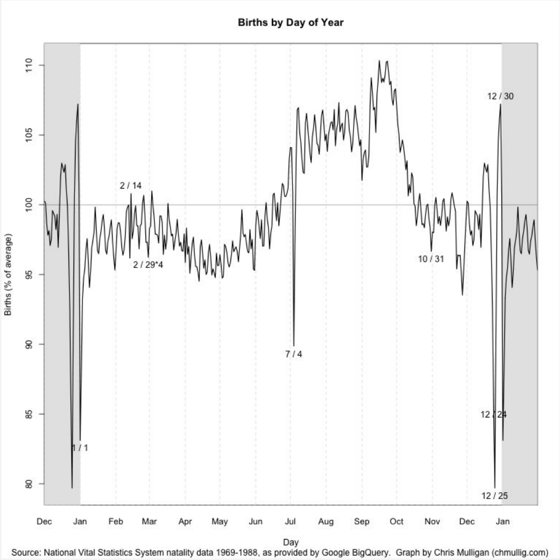
$$P(A) = 1 - P(A^c) = 1 - \frac{364^k}{365^k}$$

- Find the minimum number of people k such that p(k; m, n) meets a probability threshold q: $P(A) = 1 - \frac{364^k}{365^k} = 0.5 \; -> \; \frac{364^k}{365^k} = 0.5$

$$k \times ln(\frac{364}{365}) = ln(0.5) \; -> k \approx 252.65$$

Real-Life Applications

- In reality, some birthdays tend to occur more frequently than others, as displayed by the graph from Chris Mulligan:



- The birth rate on average increases during the first half of the year, reaching its peak around September, before declining.
- There are a few interesting spikes as well as drops on specific days as well

Appendix: colMatches Function

In The colMatches function, which is an 18.05 function, within the context of birthday probability, counts the number of shared birthdays or matches within a group of individuals.

Simulating at least 2 people sharing the same birthday (*m* = *2*)

```
# Setting up parameters
source("colMatches.r")
days = 365  # Total number of days in a year
people = 25  # Number of people in each trial
trials = 10000  # Number of simulation trials
sizematch = 2  # Desired size of the match (at least two people
sharing a birthday)

year = 1:days  # Days in a year

# Generate random birthdays for all trials
y = sample(year, people * trials, replace = TRUE)
trials = matrix(y, row = people, col = trials)

# Use colMatches function to count matches of size sizematch
within each trial
matches = colMatches(trials, sizematch)

# Calculate the probability of having at least sizematch people
sharing a birthday within each trial
prob_match = mean(matches)
```

To simulate at least 3 people sharing the same birthday (*m* = *3*), we just update sizematch to 3 for the same code.

Thank You!