Probability, Computation and Simulation Homework 4

Brian Cervantes Alvarez

October 22, 2024

---

title: "Probability, Computation and Simulation Homework 4" author: "Brian Cervantes Alvarez" date: October 22, 2024

Problem

I've read about the birthday problem, and how you only need 23 randomly chosen people for there to be a 50% chance that two people share a birthday. But how many people would you need for there to be a 50% chance that every possible birthday is represented by at least one person?

To solve this, we will estimate the probability that all birthdays (excluding February 29th and assuming each day is equally likely) are covered with M people for various values of M, identifying the value where the probability is 0.5.

Steps:

1. **Write a function to estimate** $p_M = P(\textbf{All birthdays are represented with } M \textbf{ people})$ using simulations.
2. **Try different values for** M to find where p_M is approximately 0.2 and 0.8.
3. **Construct a data frame** with a range of M values between those found in step 2.
4. **Add a new column** $\hat{p}$ with the estimated probabilities.
5. **Create a plot of** M **versus the estimated probabilities**, including confidence intervals based on the Central Limit Theorem (CLT).
6. **Find the value of** M that satisfies $p_M = 0.5$.
7. **Generalize steps 2–6** with different simulation numbers $B = 10000, 50000, 100000$.
8. **Compare the results** from different values of B and explain the findings.

Solution

1. Writing the Function $p_{\hat{M}}$

We define a function $p_{\hat{M}}$ that estimates the probability that all 365 birthdays are represented among M people:

```r
p_hat <- function(n = 365, M = 3000, B = 10000, prob = rep(1, n) / n) {
  successes <- replicate(B, {
    birthdays <- sample(1:n, M, replace = TRUE, prob = prob)
    length(unique(birthdays)) == n
  })
  mean(successes)
}
```

This function:

- Simulates B groups of M people.
- For each group, samples M birthdays with equal probability.
- Checks if all $n = 365$ birthdays are represented.
- Returns the estimated probability $\hat{p}$.

2. Finding M Values for Approximate Probabilities 0.2 and 0.8

We iteratively test different M values to find where $\hat{p}$ is approximately 0.2 and 0.8.

- Starting with $M = 1200$:
  - $p_{\hat{M}}(M = 1200) \approx 0.2$
- Testing $M = 1800$:
  - $p_{\hat{M}}(M = 1800) \approx 0.8$

3. Constructing the Data Frame

We create a sequence of M values between 1200 and 1800:

```r
M_values <- seq(1200, 1800, by = 50)
```

4. Estimating Probabilities for Each M

For each M in `M_values`, we estimate $\hat{p}$: