

Fork Selection Simulation

Luke Cadagin

11/23/2021

Load necessary packages:

```
library(tidyverse)
```

Set seed for simulation:

```
set.seed(160)
```

I have 12 forks in my drawer. 2 are of type A and 10 are of type B. In the past week I have selected a type A fork 3 out of 7 times for dinner. I would like to know the probability of this occurring to judge if there is some sort of bias associated with my fork selection (I subconsciously prefer one fork type over the other, or the forks are not shaken up enough to allow for random selection).

First we simulate a single week (7 days) of selecting a fork at dinner. We will use a binomial distribution to model this data (1 = Select Fork Type A, 0 = Select Fork Type B):

```
days_week <- 7
```

```
fork_selection_sample_week <- rbinom(n = days_week, size = 1, prob = (2/12))
```

Below is the result of our simulation:

```
fork_selection_sample_week
```

```
## [1] 0 0 0 0 0 0 0
```

Now we generate 1000 of these week simulations using the map() function:

```
fork_selection_sample_1000_weeks <- map(1:1000, ~rbinom(n = days_week, size = 1, prob = (2/12)))
```

Below are the top 6 results of our simulation:

```
head(fork_selection_sample_1000_weeks)
```

```
## [[1]]
## [1] 1 0 0 1 1 0 0
##
## [[2]]
## [1] 0 0 0 0 0 0 0
##
## [[3]]
## [1] 0 1 0 0 0 1 0
##
## [[4]]
## [1] 0 0 0 1 1 0 0
##
## [[5]]
## [1] 0 0 0 0 0 0 1
```

```
##
## [[6]]
## [1] 1 0 1 0 0 0 0
```

Now we calculate the proportion of Type A forks selected in each sample using the `map()` function again:

```
prop_fork_selection_sample_1000_weeks <- map(fork_selection_sample_1000_weeks, ~sum(.) / length(.))
```

Below are the top 6 proportions of our simulation:

```
head(prop_fork_selection_sample_1000_weeks)
```

```
## [[1]]
## [1] 0.4285714
##
## [[2]]
## [1] 0
##
## [[3]]
## [1] 0.2857143
##
## [[4]]
## [1] 0.2857143
##
## [[5]]
## [1] 0.1428571
##
## [[6]]
## [1] 0.2857143
```

Next we unlist this data:

```
head(unlist(prop_fork_selection_sample_1000_weeks))
```

```
## [1] 0.4285714 0.0000000 0.2857143 0.2857143 0.1428571 0.2857143
```

Finally, we calculate the proportion of simulated samples where fork A was selected greater than $(3/7)$ of the time:

```
tibble(prop = prop_fork_selection_sample_1000_weeks) %>%
  filter(prop >= (3/7)) %>%
  summarize(prop = n() / 1000)
```

```
## # A tibble: 1 x 1
##   prop
##   <dbl>
## 1 0.094
```

We see that the proportion of simulated samples where fork A was selected greater than $(3/7)$ of the time is .094.

.094 is above an alpha of .05. This leads me to assert that the outcome of my fork selection last week was highly unlikely, yet not statistically significant. I would need to gather more data about my fork selection to determine if there is any bias in my selection of forks at dinner time.

We can also calculate this probability using the `pbinom()` function:

```
1 - pbinom(2,7,(2/12))
```

```
## [1] 0.09577546
```