Methods Camp 2025: Day 3

Christina Pao, Sofia Avila, Florencia Torche

September 6, 2025

Table of contents

Sampling Functions

SOC 500 Precursor

Understanding social statistics is very based on understanding randomness and random processes!

A **random variable** is a variable whose value is determined by the outcome of a random phenomenon.

- **Discrete Random Variables:** Can take on a countable number of values (e.g., coin flips, die rolls)
- **Continuous Random Variables:** Can take any value within a range (e.g., height, temperature)

Basic Example: Coin Flips

Open your `day3_afternoon.qmd` file to code along and annotate!

```r
# Simulating a single coin flip
coin_flip <- sample(c("Heads", "Tails"), size = 1)
print(coin_flip)
```

```
[1] "Heads"
```

```r
# Multiple coin flips
n_flips <- 10
coin_flips <- sample(c("Heads", "Tails"), size = n_flips, replace = TRUE)
print(coin_flips)
```

```
 [1] "Heads" "Tails" "Heads" "Heads" "Tails" "Heads" "Tails" "Heads" "Heads"
[10] "Tails"
```

```r
# Converting to numeric (1 for Heads, 0 for Tails)
numeric_flips <- ifelse(coin_flips == "Heads", 1, 0)
print(numeric_flips)
```

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

Simulating multiple flips: `sample()`

Task: Simulate rolling a six-sided die 20 times and calculate the frequency of each outcome.

```r
die_rolls <- sample(1:6, size = 20, replace = TRUE)
table(die_rolls)
```

```
die_rolls
1 2 3 4 5 6
1 3 4 2 4 6
```

Probability Distribution Functions and Their Meanings

Function	Description
`d____( )`	Density/probability function (PDF/PMF)
`p____( )`	Cumulative distribution function (CDF)
`q____( )`	Quantile function (inverse CDF)
`r____( )`	Random number generation

Let’s draw these on the board!

Two ways to get an answer[1]

- Analytic solutions
  - exact answers
  - expressed as mathematical formulas with variables
  - limited in scope (for problems that can be solved mathematically)
- Simulation solutions
  - approximate
  - computational
  - resource intensive
  - easier for many complex systems!

You will use both in SOC 500!

Binomial Distribution: Flipping a (Fair) Coin

Scenario: Flip a fair coin 10 times, counting heads. What is the probability of getting exactly 7 heads in 10 flips?

Analytic Solution

Probability of getting exactly 7 heads in 10 flips:

$$P(X = 7) = \binom{10}{7} \cdot (0.5)^7 \cdot (0.5)^3 = \binom{10}{7} \cdot (0.5)^{10}$$

Analytic Solution in R

```r
# Probability of getting exactly 7 heads in 10 flips
n <- 10
p <- 0.5
k <- 7
```

```r
# Using dbinom for exact probability
prob_7_heads <- dbinom(x = k, size = n, prob = p)
prob_7_heads
```

```
[1] 0.1171875
```

What about 7 or fewer heads?

Probability of getting 7 or fewer heads:

$$P(X \leq 7) = \sum_{i=0}^{7} \binom{10}{i} \cdot (0.5)^i \cdot (0.5)^{10-i} = \sum_{i=0}^{7} \binom{10}{i} \cdot (0.5)^{1}$$

```r
prob_7_or_less <- pbinom(q = k, size = n, prob = p)
prob_7_or_less
```

```
[1] 0.9453125
```

Simulation Solution

```r
# Simulate many experiments
n_simulations <- 10000
results <- rbinom(n = n_simulations, size = 10, prob = 0.5)
```

```r
# Estimate probability of getting exactly 7 heads
prob_7_heads_sim <- mean(results == 7)
round(prob_7_heads_sim, 4)
```

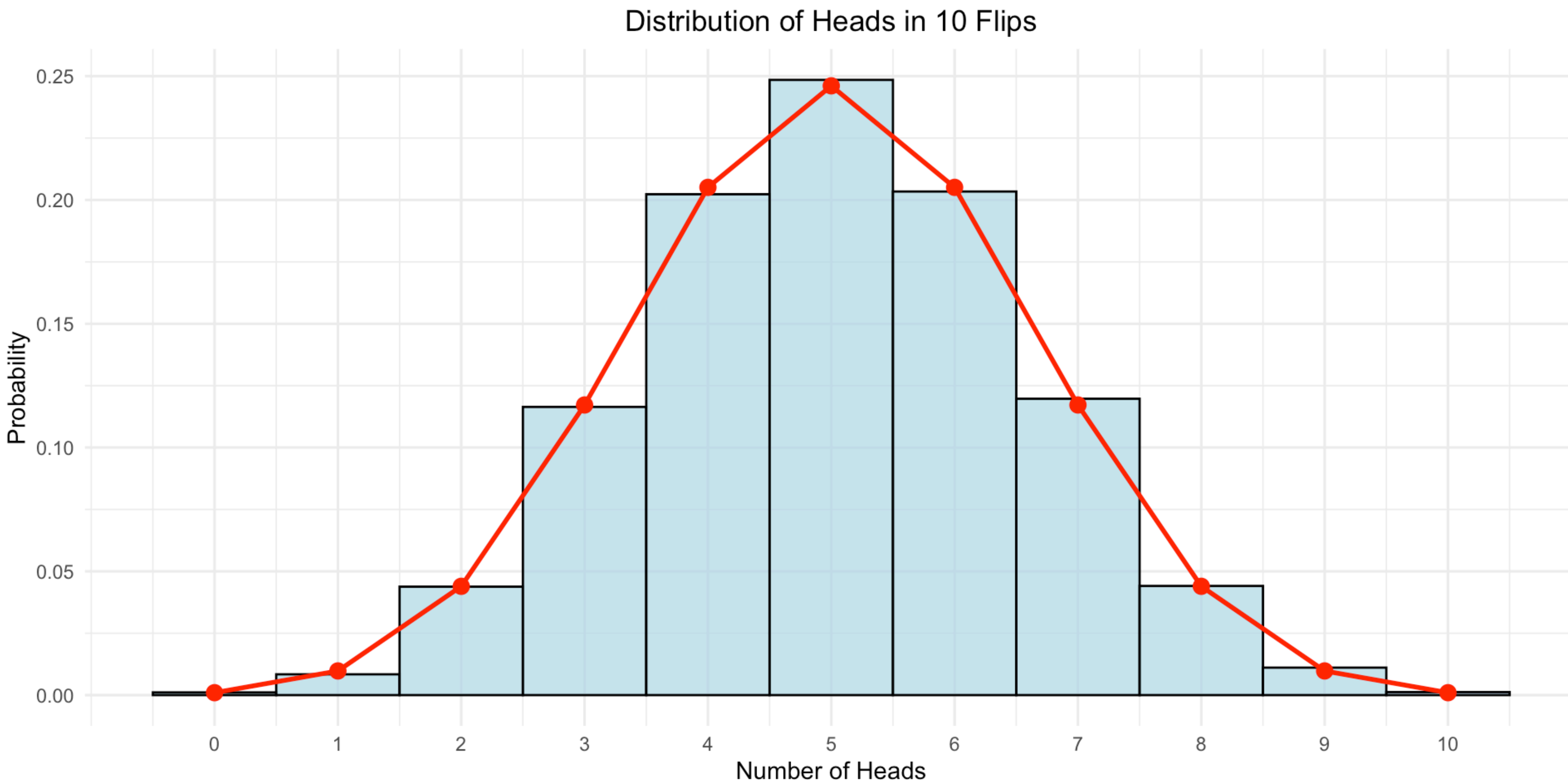
```
[1] 0.1197
```

```r
# Create data frame for visualization
sim_data <- data.frame(heads = results)
theoretical_data <- data.frame(
  heads = 0:10,
  probability = dbinom(0:10, size = 10, prob = 0.5)
)
```

Plotting the different solutions

```r
# Visualize the distribution
library(ggplot2)

ggplot() +
  geom_histogram(data = sim_data, aes(x = heads, y = after_stat(density)),
                 binwidth = 1, center = 0, fill = "lightblue",
                 color = "black", alpha = 0.7) +
  geom_point(data = theoretical_data, aes(x = heads, y = probability),
             color = "red", size = 3) +
  geom_line(data = theoretical_data, aes(x = heads, y = probability),
            color = "red", size = 1) +
  labs(title = "Distribution of Heads in 10 Flips",
       x = "Number of Heads",
       y = "Probability") +
  scale_x_continuous(breaks = 0:10) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```


Distribution of Heads in 10 Flips
0.25
0.20
0.15
0.10
0.05
0.00
Probability
0
1
2
3
4
5
6
7
8
9
10
Number of Heads

Your turn!

We are now going to use the normal distribution instead of the binomial distribution (`rnorm()`, `pnorm()`, `dnorm()`, `qnorm()`).

Scenario: Let's say the heights of adults are a mean of 170 cm and a standard deviation of 10 cm.

1. What proportion of adults are taller than 180 cm?
2. What is the 90 percentile height? (i.e., what is the height that 90% of adults are shorter than?)

Use the scaffolding for both the analytic and simulation solutions!

Pair work!

Assignment

We are now going to have you work in new groups to combine what you've learned today about making/applying functions and random sampling.

Any questions?