

# CompSci 190: Sampling

Jeff Forbes

October 22, 2018

According to a [survey](#) conducted in August and made public on the Brookings Institution website, a plurality of college students polled, 44 percent, believed that hate speech was not protected by the First Amendment.

# Probability Distribution

---

- Random quantity with various possible values
  - “Probability distribution”:
    - All the possible values of the quantity
    - The probability of each of those values
  - In some cases, the probability distribution can be worked out mathematically without ever generating (or simulating) the random quantity
-

# Empirical Distribution

---

- Based on observations
  - Observations can be from repetitions of an experiment
  - “Empirical Distribution”
    - All observed values
    - The proportion of counts of each value
-

# Multiplication Rule

---

Chance that two events  $A$  and  $B$  both happen

=  $P(A \text{ happens}) \times P(B \text{ happens given that } A \text{ has happened})$

- The answer is *less than or equal to* each of the two chances being multiplied
  - The more conditions you have to satisfy, the less likely you are to satisfy them all
-

# Addition Rule

---

If event  $A$  can happen in *exactly one* of two ways, then

$$P(A) = P(\text{first way}) + P(\text{second way})$$

- The answer is *greater than or equal to* the chance of each individual way
-

# Example: At Least One Head

---

- In 3 tosses:
  - Any outcome *except* TTT
  - $P(\overline{\text{TTT}}) = (1/2) \times (1/2) \times (1/2) = 1/8$
  - $P(\text{at least one head}) = 1 - P(\overline{\text{TTT}}) = 7/8 = 87.5\%$

- In 10 tosses:
  - $1 - (1/2)^{10} = 99.9\%$

<http://bit.ly/FoDS-f18-1022>

- Roll a fair die 6 times.
    - What is  $P(\text{get at least one 6})$ ?
-

# Sampling

---

- Deterministic sample:
  - Sampling scheme doesn't involve chance
- Probability sample:
  - Before the sample is drawn, you have to know the selection probability of every group of people in the population
  - Not all individuals have to have equal chance of being selected
- Uniform Random Sample: ?

(Demo)

---



# Sample of Convenience

---

- Example: sample consists of whoever walks by
  - Just because you think you're sampling "at random", doesn't mean you are.
  - If you can't figure out ahead of time
    - what's the population
    - what's the chance of selection, for each group in the populationthen you don't have a random sample
-

# Law of Averages

---

If a chance experiment is repeated many times, independently and under the same conditions, then the proportion of times that an event occurs gets closer to the theoretical probability of the event

As you increase the number of rolls of a die, the proportion of times you see the face with five spots gets closer to  $1/6$

(Demo)

---

# Large Random Samples

---

If the sample size is **large**,

then the **empirical distribution** of a *uniform random sample*

resembles the **distribution of the population**,

with **high probability**

---

# Final Project

---

- Demonstrate proficiency in techniques from class applied to dataset **of your choosing**
  - Data!
    - Domain: ?
    - Scale: Large enough
      - Categorical and numerical variables
      - More than 100 observations
    - Accessible & Manageable
-