

Data 8, Lab 5

Probability and Sampling

Hubert Luo

Fall 2019

4 October 2019

Announcements

- Project 1 due tonight
- Midterm is in 2 weeks (10/18)

Agenda

1. Probability
2. Sampling
3. Worksheet (Iteration and Conditionals)
4. Lab Notebook

Probability Review

- Probability is a number between 0 and 1 (inclusive)
- $P(A) = 0.5$ means there is a 50% chance of event A happening
- Calculated by determining what % of all possible outcomes result in A happening
- Complement: $P(A \text{ doesn't happen}) = 1 - P(A)$
- Multiplicative Rule: Probability that events A and B happen is $P(A \text{ happens}) \times P(B \text{ happens given that A has happened})$
- Additive Rule: If event A can happen in two ways, then $P(A \text{ happens}) = P(\text{first way}) + P(\text{second way})$

Probability Questions

A standard deck of cards is 52 cards, 13 of each suite. None of the questions depend on the previous question. All draws are without replacement. Keep all solutions as fractions!!

1. What is the probability of drawing an ace?
2. What is the probability of drawing two queens in a row?
3. What is the probability of getting both a king and a queen after two draws?
4. What is the probability of getting at least one ace when dealt 5 cards?

Probability Questions: Solutions

1. What is the probability of drawing an ace?

$$4/52 = 1/13$$

2. What is the probability of drawing two queens in a row?

$$\text{Multiplicative Rule: } 4/52 \times 3/51$$

3. What is the probability of getting both a king and a queen after two draws?

$$\text{Additive Rule: } 4/52 \times 4/51 + 4/52 \times 4/51$$

4. What is the probability of getting at least one ace when dealt 5 cards?

$$1 - P(\text{No Aces}) = 1 - (48/52 \times 47/51 \times 46/50 \times 45/49 \times 44/48)$$

Sampling

- Random sample: Know the probability every individual in the population is in the sample (doesn't have to be equal)
- Empirical Distribution: Based on observed values, i.e., from a simulated experiment
- Probability Distribution: Based on theoretical calculations
- Law of Large Numbers: Repeating an experiment a large number of times will cause the empirical probability of an event to approach its theoretical probability
 - Example: The more times you flip a coin, the closer the proportion of heads gets to 0.5 (usually)

Inference

- Extrapolate information about a sample to make conclusions for the population
- Population parameter: A metric associated with a population
- Sample statistic: A metric associated with a sample
 - The theoretical distribution of a statistic can be difficult to determine, so we often approximate it using its empirical distribution
 - Good approximation if the number of samples is large

Lab Notebook