DSC 10, Spring 2018
Lecture 13

Probability, Sampling, Statistics

sites.google.com/eng.ucsd.edu/dsc-10-spring-2018

Credit: Anindita Adhikari and John DeNero

Announcements

Midterm on Wednesday

- Practice exams on class website
- Written exam
  - “paper coding”
  - No notes, books, phones, computers
  - Will be given a reference sheet (also on website)
- Covers through last week (Monty Hall)
- Check Piazza for seat assignment **before** coming to the exam

Homework Submission Reminders

- When submitting to Gradescope
  - Submit only one assignment per group; add your partner to your submission
  - Mark which page each question appears on

DSC Town Hall Meeting Tonight

- 5-7pm tonight
- Atkinson Hall Auditorium
- Ask questions
- Get updates on future plans for the major

Probability

Probability

- Event: *some* of the possible outcomes
- Lowest value: 0
  - Chance of event that is impossible
- Highest value: 1 (or 100%)
  - Chance of event that is certain

Equally Likely Outcomes

Assuming all outcomes are equally likely, the chance of an event A is:

$$P(A) = \frac{\text{number of outcomes that make A happen}}{\text{total number of outcomes}}$$

Discussion Question

I have three cards: red, blue and green.

What is the chance that I choose a card at random and it is green, then without putting it back, I choose another random card and it is red?

A. 1/9
B. 1/6
C. 1/3
D. 2/3
E. None of the above

Multiplication Rule

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

= P(*A* happens) x P(*B* happens given that *A* 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

Discussion Question

I have three cards: red, blue and green. I pick one card, then without putting it back, I pick a second card. What is the probability that I pick one red and one green card?

A. 1/6
B. 1/3
C. 5/6
D. None of the above

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
- The more different ways an event can happen, the more likely it is to occur

Discussion: At Least One Head

I have a fair coin.

Find the probability of at least one head in 3 tosses.

Discussion: At Least One Head

I have a fair coin.

Find the probability of at least one head in 3 tosses.

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

Discussion Question

Every time I call my Grandma, the probability that she answers her phone is ⅓. If I call my Grandma **two** times today, what is the chance that I will talk to her?

A. 1/3
B. 2/3
C. 1/2
D. 1
E. None of the above

Discussion Question

Every time I call my Grandma, the probability that she answers her phone is ⅓. If I call my Grandma **three** times today, what is the chance that I will talk to her?

A. 1/3
B. 2/3
C. 1/2
D. 1
E. None of the above

Sampling

Sampling

- Deterministic sample:
  - Sampling scheme doesn’t involve chance
- Probability sample:
  - Before the sample is drawn, you have to know the probability of selecting each group of people in the population
  - Not all individuals need to have an equal chance of being selected

A probability sample

- Population: 3 individuals (A, B, C)
- Select a sample of 2 as follows:
  - A chosen with probability 1
  - Choose B or C based on coin toss
- Possible pairs: AB, AC, BC
  - Chance of AB: ½
  - Chance of AC: ½
  - Chance of BC = 0

(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 population

  then you don't have a random sample

Distributions

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

(Demo)

Large Random Samples

Law of Averages

If a chance experiment is repeated

- many times,
- independently,
- under the same conditions,

then the proportion of times that an event occurs gets closer to the theoretical probability of the event.

Ex. As you roll a die repeatedly, the proportion of times you roll a 5 gets closer to $\frac{1}{6}$.

Large Random Samples

If the sample size is large,

then the empirical distribution of a uniform random sample

matches the distribution of the population,

with high probability.

(Demo)

At Least One Six

If you roll a die 4 times, what is the probability of getting at least one 6?

A. $\frac{5}{6}$
B. $1 - \frac{5}{6}$
C. $1 - (\frac{5}{6})^4$
D. $1 - (\frac{1}{6})^4$
E. None of the above.

What's the general formula, if you roll a die *n* times?

Statistics

Why sample?

Probability

Statistics

Sampling

Estimation

Statistical Inference:

Making conclusions based on data in random samples

Example:

fixed

Use the data to guess the value of an unknown number

depends on the random sample

Create an **estimate** of the unknown quantity

Terminology

Parameter

A number associated with the population

Statistic

A number calculated from the sample

A statistic can be used as an **estimate** of a parameter