DATA 8

Fall 2016

Lecture 14, September 26

Probability and Sampling

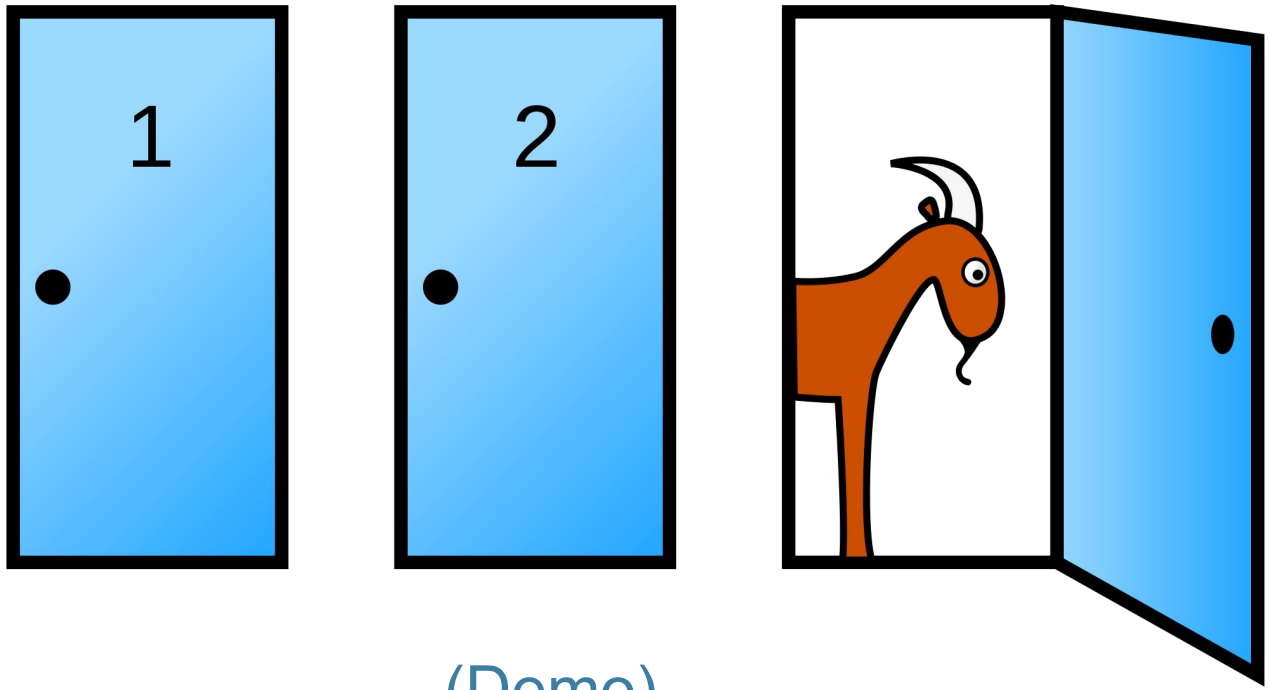
Slides created by Ani Adhikari and John DeNero

Announcements

- Project 1 Parts 1 and 2 Checkpoint: 5 pm tomorrow Tuesday 9/27.

- No homework due this week. Homework will be assigned on Friday.

Monty Hall Problem



(Demo)

Another Way to See the Answer

Contestant Picks	Monty Throws Out	Remaining Door
Car	Goat	Goat
Goat	Goat	Car
Goat	Goat	Car

Only this column is random.
Each row has chance 1/3.

Probability

- Lowest value: 0
  - Chance of event that is impossible
- Highest value: 1 (or 100%)
  - Chance of event that is certain

- If an event has chance 70%, then the chance that it doesn't happen is
  - 100% - 70% = 30%
  - 1 - 0.7 = 0.3

(Demo)

Equally Likely Outcomes

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

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

(Demo)

Fraction of a Fraction

(Demo)

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 *smaller* than each of the two chances being multiplied.
- The more conditions you have to satisfy, the less likely you are to satisfy them all.

(Demo)

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 *bigger* than each the chance of each individual way.

At Least one Head

- In 3 tosses:
  - Any outcome *except* TTT
  - P(TTT) = (½) x (½) x (½) = ⅛
  - P(at least one head) = 1 - P(TTT) = ⅞ = 87.5%

- In 10 tosses:
  - 1 - (½)**10
  - 99.9%

(Demo)

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.

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.