DATA 8

Fall 2016

Lecture 15, September 28

Empirical Distributions

Slides created by Ani Adhikari and John DeNero

Announcements

- There is lab today and tomorrow.
  - Probability discussion worksheet
  - Project time (due 5 pm Tuesday Oct 4)
- No homework due this week. Homework will be assigned on Friday.
- Midterm is on Friday Oct 14, two weeks away. No computers or calculators on the midterm.

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.

(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.

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 many cases, the probability distribution can be worked out mathematically without ever generating the random quantity

(Demo)

Empirical Distribution

- Based on observations
- Observations are typically from repetitions of an experiment
- “Empirical Distribution”
  - All observed values
  - The proportion of repetitions that produced each value

(Demo)

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 random sample

resembles the distribution of the population,

with high probability.

Roulette


MIN
MAX
CURRENT
INSIDE
OUTSIDE
Amount to bet per click
$1
.25
1
5
25
Clear
Repeat
Spin
1st 12
2nd 12
3rd 12
1 - 18
Even
Odd
19-36
2 to 1
2 to 1
2 to 1