



190F
Fall 2018

Foundations of Data Science

Lecture 13

Sampling

Announcements

Netflix Is No. 1 Company Techies Want to Work For, Survey Finds

Variety • 2 hours ago



Netflix took the No. 1 spot on Hired's 2018 survey of tech workers ranking the companies they'd most like to work for, moving up from fifth place on the job site's 2017 survey. That put the streamer ahead of Google (#2), Elon Musk's Tesla (#3) and SpaceX (#4), Airbnb (#5) and Apple (#6). The survey polled 2,200 tech workers in the U.S., Canada and the U.K.

Sampling

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

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

(Demo)

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 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

Probability & Simulation

Calculation

Roll a fair die 4 times.

What is $P(\text{get at least one } 6)$?

(Demo)
