Lecture 11: Hypothesis testing

COSC 480 Data Science, Spring 2017

Michael Hay

Example

- A Ferengi government official shows you census data asserting that in 10,000 families surveyed, there are just 5,000 girls.

  (Recall that E[G] = 1.)

Commander *Data* of Star Trek

- Has the data been tampered? Are Ferengi becoming less sexist?

http://www.fanpop.com/clubs/lt-commander-data/images/381182/title/sherlock-data-photo

Example

- British biologist and statistician Ronald Fisher met a lady who claimed upon tasting her tea, she could tell whether the tea or milk was added to the cup first.

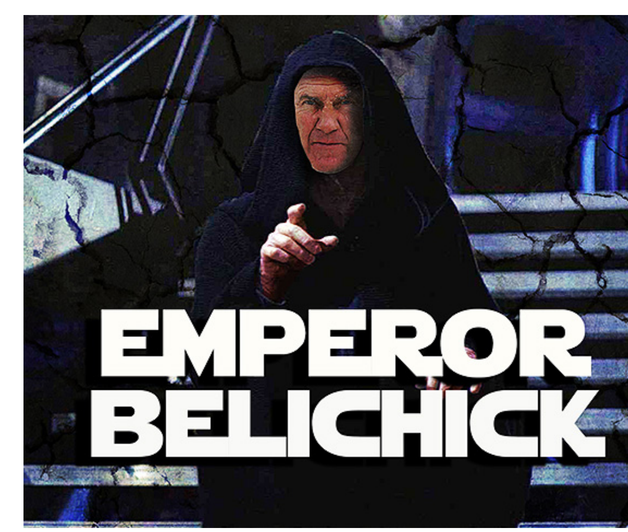
- Fisher gave her eight cups, four of each kind. The cups were ordered randomly. And the lady was asked to identify the 4 where milk was added first.

- The lady correctly chooses all 4. Did she get lucky?

Example

- The NFL is supposed to use only fair coins for the coin toss used in football games, both at the start of the game and the start of overtime. (Due to manufacturing, it's impossible to get a perfectly fair coin, but it should be close.)

- The patriots have never lost the coin toss in an overtime. Coincidence? Or



Statistical inference

Statistical inference: let X be a random variable from a distribution F. We don't know F. Given that we observe an outcome, denoted $X = x$, can we infer (some property of) F?

Hypothesis testing

- We might have a theory about *F*. Can we prove our theory is correct?

- We want to prove a hypothesis H_A but its hard so we try **disprove a null hypothesis H_0**.

- Some hypotheses are about the value of certain parameters that define *F*.

Hypotheses

- Coin flip?

- Ferengi?

- Lady tasting tea?

Test statistics

- A **test statistic** is some measurement we can make on the data which is likely to be **big under H_A** and **small under H_0**.

Hypothesis tests

- Coin flip?

- Ferengi?

- Lady tasting tea?

Errors

		Decision	
		Retain the null	**Reject the null**
Truth in the population	True	CORRECT $1 - \alpha$	TYPE I ERROR α
	False	TYPE II ERROR β	CORRECT $1 - \beta$ POWER

Exercise

Instructions: ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you

- Recall that we have census data that says among 10,000 families, there are 5,000 girls. The null hypothesis is that the data comes from a population of 10,000 Ferengi families.

- Suppose that you have a function `make_ferengi_family()` that will sample one family according to the Ferengi rules of producing offspring.

- How would you use this to estimate P_{H0}(test says "reject")?

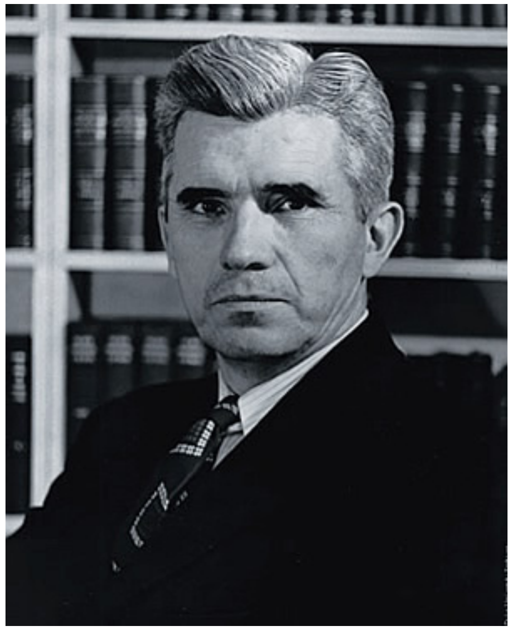
Exercise

Instructions: ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you

- Consider the lady tasting tea example. Suppose the null hypothesis is true. What is the probability that she correctly guesses which 4 out of 8 cups had the milk added first?
- Hint: if the null hypothesis is true, she is guessing randomly? How many different ways can you choose 4 cups from 8?
- You can answer using math or by describing a program you could write to estimate/calculate this.

Instructions: ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

Rhine Paradox



- Joseph Rhine was a parapsychologist in the 1950's (founder of the Journal of Parapsychology and the Parapsychological Society, an affiliate of the AAAS).
- He ran an experiment where subjects had to guess whether 10 hidden cards were red or blue.
- He found that about 1 person in 1000 had ESP, i.e. they could guess the color of all 10 cards.
- Q: what's wrong with his conclusion?

Beware of the **multiple testing problem**! (See "p hacking" in book.)

Continuous random variables

- Continuous r.v.'s take on values from some continuous range.

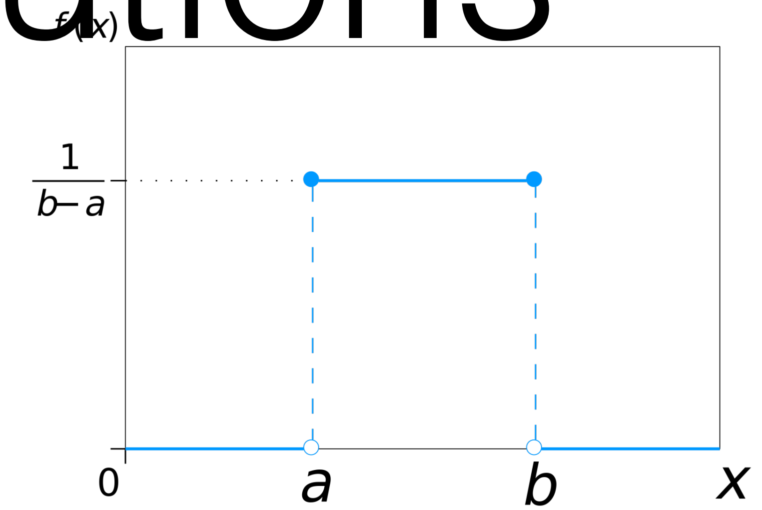
- Distribution defined by a *probability density function*, denoted *f(x).*

- Note: f(x) can be greater than 1!

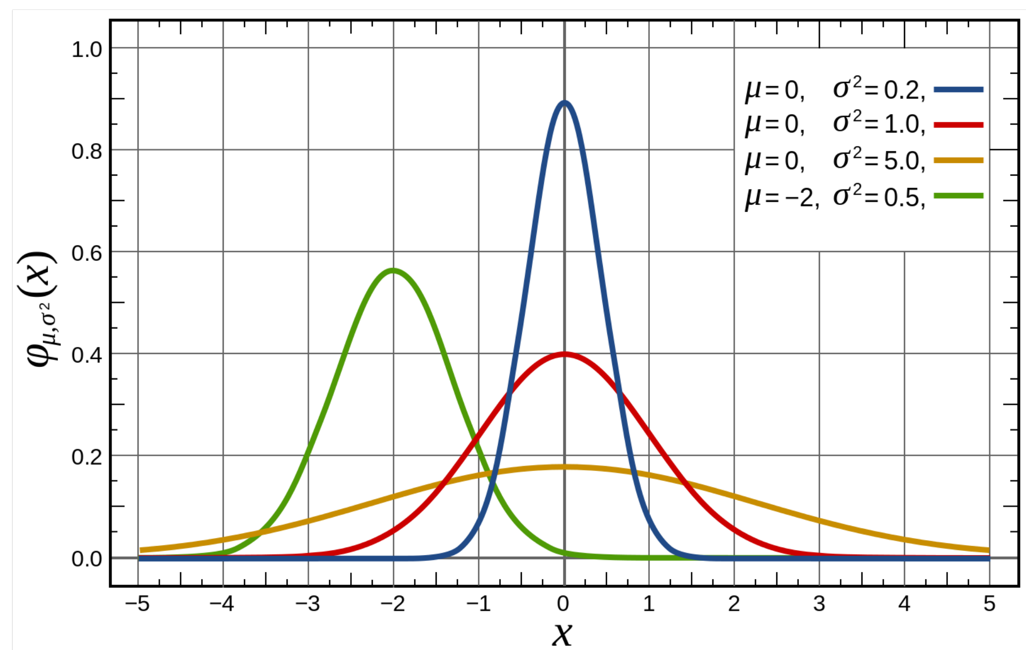
- To calculate Pr[a ≤ X ≤ b], you integrate f(x) over interval [a,b]

Example distributions

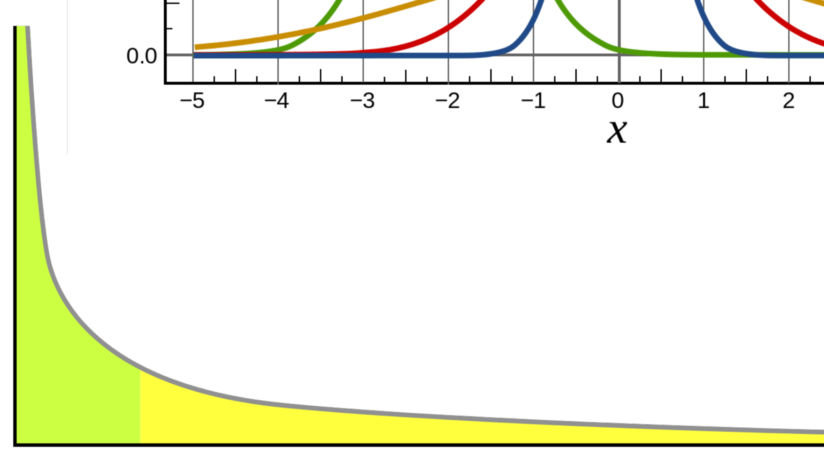
- ***Uniform*** distribution
  - E.g., outcome of a dice, `rand()`



- ***Normal*** distributions
  - aka *Gaussian*, *bell curve*
  - E.g., test scores, people's blood pressure



- ***Power-law*** distributions
  - "80/20" rule
  - E.g., income in US, # page views, # friends



http://en.wikipedia.org/wiki/File:Uniform_Distribution_PDF_SVG.svg
http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg
http://en.wikipedia.org/wiki/File:Long_tail.svg

Central limit theorem

- The distribution of the sum (or mean) of a set of n identically-distributed random variables X_i approaches a normal distribution as n goes to infinity.

- Many common parametric statistical tests, like t-test and ANOVA *assume normally-distributed data*, but depend on sample mean and variance measures of the data.

- They typically work reasonably well for data that are not normally distributed as long as the samples are not too small.

- Check your data! Use box-plot, histogram.

Important tests

- Many parametric tests assume that the data is normally distributed and that samples independent and all have the same distribution.
- T-test: compare two groups, or two interventions on one group
- Chi-squared: compare counts in a contingency table (e.g., admissions vs. gender).
- ANOVA: compare outcomes under several discrete interventions.