Data 8, Lab 10

Inference Review: Hypothesis Testing, A/B Testing, and Confidence Intervals

Hubert Luo

Fall 2019

15 November 2019

Berkeley
UNIVERSITY OF CALIFORNIA

Agenda

1. Hypothesis Testing
2. A/B Testing
3. Inference Comparison
4. Confidence Intervals

Hypothesis Testing

- Choose a **test statistic** to measure discrepancy between model and observed data
- Identify **null and alternative hypotheses**
- **Simulate** assuming the null is true and generate an array of test statistics
- **P-value** is the proportion of simulated test statistics equal to or greater than the observed test statistic

A/B Testing

- **Null Hypothesis:** The two distributions are from the same underlying distribution
- **Alternative Hypothesis:** The two distributions are **not** from the same underlying distribution (or some variation of this statement)
- **Test statistic** is difference between the metric of two distributions
- Shuffle the **labels** of the data without replacement to get an array of simulated test statistics
  - This keeps same proportion of each label!
- **P-value** is the proportion of simulated test statistics equal to or greater than the observed test statistic

A/B Testing vs. TVD

- A/B Testing
  - Used to test if two observed distributions from the same underlying distribution?
- Total Variation Distance (TVD)
  - A common test statistic used during hypothesis testing
  - Compute distance between two different samples from the same distribution

Hypothesis vs A/B Testing

- Hypothesis Testing: Is model consistent with observed data?
  - Simulate data under the null using something like **np.random.choice** or **sample_proportions**
- A/B Testing: Are two observed distributions from the same underlying distribution?
  - Shuffle the **labels** of the observed data

Bootstrapping vs Testing

- Different from hypothesis testing and A/B testing!
- Goal is to estimate a **population parameter** using simulated **sample statistics** and evaluate variability of those estimates
- If the original sample is large and selected at random, it likely resembles the population
- Instead of getting entirely new samples from the population, we resample from the original sample the same number of individuals with replacement
  - If we sample without replacement, we will always get the sample original sample back!

Confidence Intervals

- For a 95% confidence interval, it is true that:
  - Out of 100 confidence intervals, we expect 95 of them to contain the true population parameter
  - There is a 95% probability the confidence interval contains the population parameter
    - The confidence interval was random so this is true
    - If the confidence interval was fixed, like [0.5,1] this then becomes false!
- It is **not** true that there is a 95% probability a population parameter is in a confidence interval.
  - This is because the population parameter is fixed so we can't so anything probabilistic about it

Announcements

- Project 2 due tonight! (11/15)
- HW11 due next Thursday (11/21)
- Project 3 – Checkpoint 1 due next Friday (11/22)