STA130H1S TUT0109
W6: Sampling, Bootstrap Samples, and Confidence Intervals

Feb 15, 2019

Email: sonia.chhay@mail.utoronto.ca
Website: soniachhay.github.io

Announcements

- No class and tutorial next week (reading week!)
  - No OH but Piazza will be checked regularly
  - New College Stats Aid Centre will be open during reading week (Wetmore Hall, Room 68A)
- Example midterms have been posted to Quercus
  - Midterm details are on the next slide

About the Midterm

- When: During your usual tutorial time (**Fri March 1st**)
- Where: You MUST attend the correct section's midterm
  - **AM section**: EX 200
  - **PM section**:
    - MS 3154: Last names from A - Lo
    - WB 116: Last names from Lu - Z
- Includes: All material up to & including Feb 25th (mostly a review class)
- Format:
  - Multiple choice
  - Fill in the blanks
  - Written answers (**make sure to write in complete sentences**)

Overview

- Material and vocabulary review
- Group discussion
- Oral presentations

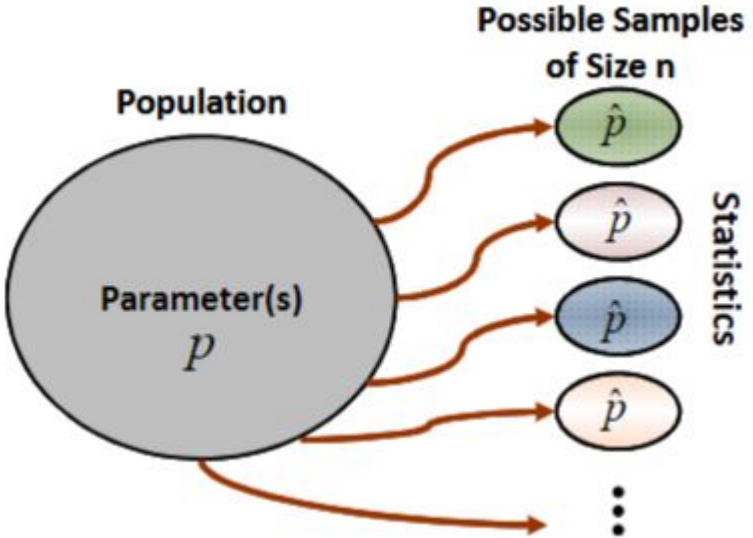
Vocabulary for this Week (1/2)

- Percentile (Quantile)
- Sampling distribution
- Population
- Sample
- Parameter
- Statistic

Percentile (Quantile)

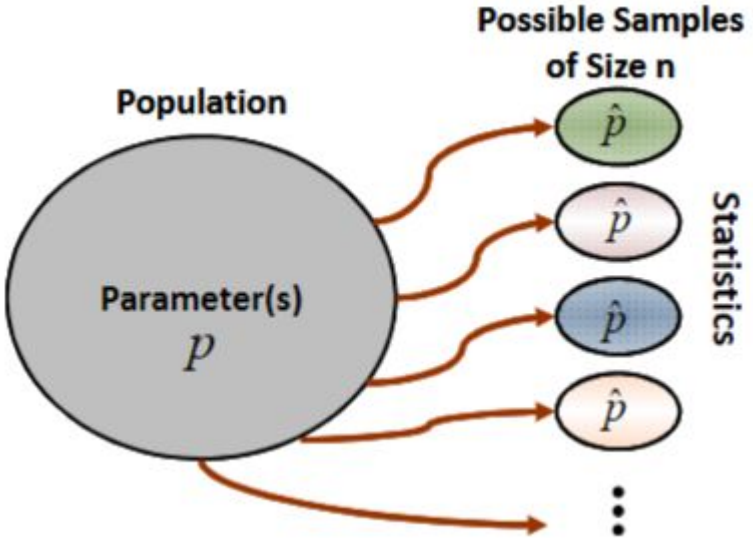
- Percentile: A measure *p* between 0 and 100, in which the p^{th} percentile is the smallest value that is larger or equal to p% of all the values
  - 25^{th} percentile = Quarter of points are less than it
- Quantile: Divides the data into equally sized groups
  - 0.25 or 25% quantile = Quarter of the points are less than it

Sampling Distribution



- Sampling distribution of a statistic: Distribution of statistic values taken for all possible samples of the same size (n) from the same population
- Sampling distribution:
  - Take random samples from a certain population
  - Calculate the statistic of each sample
  - Plot and observe distribution

Sampling Distribution (break-down)



- Population: A whole - Every member/element of a group
- Sample: A fraction or percentage of that group
- Parameter (p): Value that tells you something about a population
- Statistic (p̂): Tells you something about a small *part* of the population

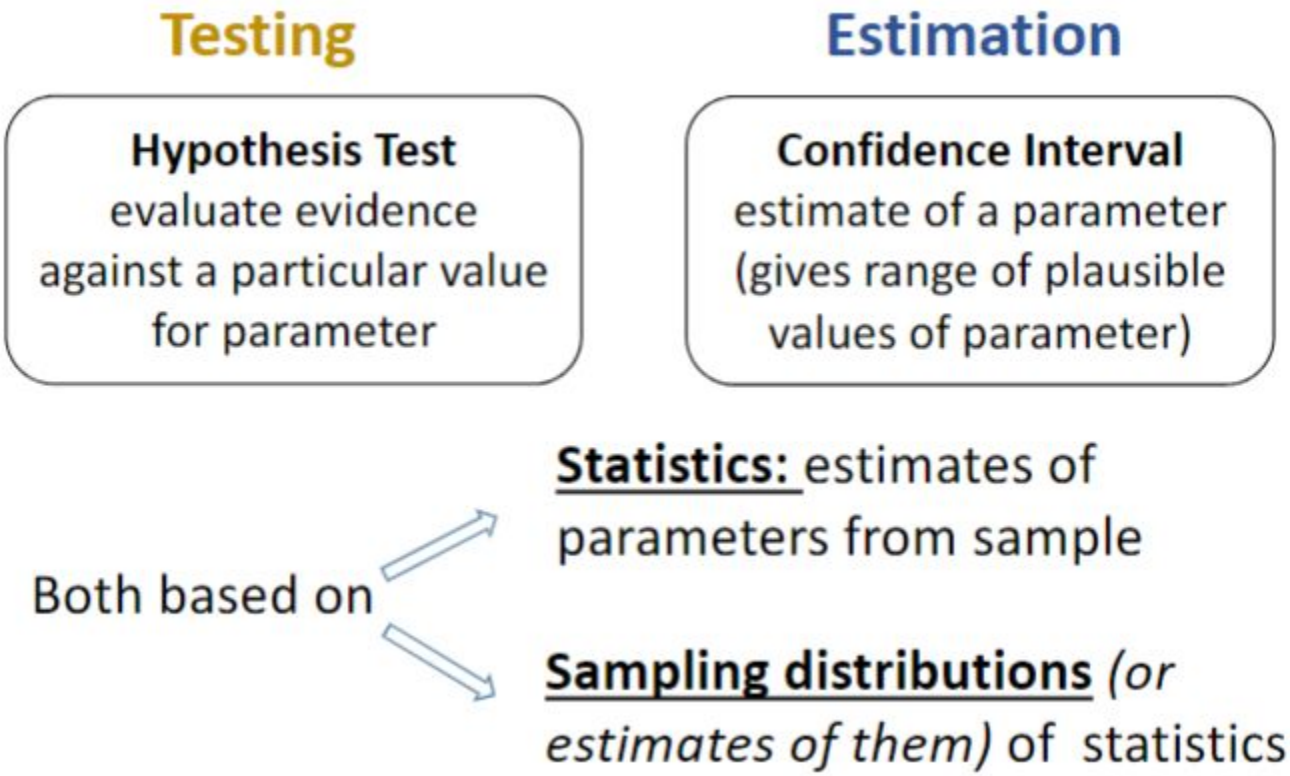
https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/what-is-sampling-distribution/v/introduction-to-sampling-distributions

Vocabulary for this Week (2/2)

- Testing
- Estimation
- Representative

- Resampling
- Bootstrap
- Confidence interval
- Confidence level

Testing and Estimation



Resampling and Bootstrap

- Bootstrap: An approach that helps us describe the variability of statistics based on one sample
  - Does not create new data, only reuses sample data!
  - Resampling method used to estimate statistics on a population by sampling a dataset with replacement
  - I.e. Resampling from the sample -> Take the sample that you have and resample from it
- Purpose of bootstrap: Estimate the sampling distribution of a statistic
  - Allows us to get a confidence interval (CI)

Confidence Interval & Confidence Level

Recall: Using only our sample data, we are trying to come up with a range of values that would be plausible for the true parameter value

- Confidence interval (CI): Captures the parameter value a certain percentage of the time
  - E.g. 95% CI includes the parameter for 95% of possible samples
  - Here, 95% is called the confidence level
- Purpose of CI: Obtain an estimate for the parameter that reflects sampling variability
- Examples of when you could use CI:
  - Wish to estimate proportion of people living in Toronto who use the TTC
  - Number of coffees people in this class drink each week

Confidence Interval (Cont'd)

- Wide CI: If you had taken a different sample from the population, you could arrive at a very different estimate
- Narrow CI: If you had taken a different sample from the population, you could expect to get a similar estimate

Note: Always check that your CI range makes sense!

- E.g. If you reported CI for a proportion, it must be bounded by [0;1]