Lecture 11: Estimation

Lecturer: Jie Fu

EEL 3850

Motivating problem

- Estimating the average height of 3rd grade students in schools

Data sample 1:

53.49	51.59	53.94	56.57	51.30	51.30
56.74	54.30	50.59	53.63	50.61	50.60
52.73	46.26	46.83	50.31	48.96	52.94
49.28	47.76	56.40	51.32	52.20	47.73

Average = 51.51 inches

Data sample 2:

52.33	48.55	53.13	50.20	51.12	50.19
57.56	51.96	48.83	54.47	48.34	52.63
46.12	48.02	52.59	54.22	52.51	51.65
51.10	47.56	49.84	50.62	55.17	53.03

Average = 52.56 inches

Classical inference

- unknown parameter θ as a deterministic (not random!) but unknown quantity.
  - Average height.
  - The fraction of voters who support this candidate

- The observation X is random and its distribution
  - $p_X(x;\theta)$ if X is discrete
  - $f_X(x;\theta)$ if X is continuous
    - depends on the value of θ.

Classical inference

- unknown parameter θ as a deterministic (not random!) but unknown quantity.
  - E.g. Average height.

- The observation X is random and its distribution
  - E.g. Average height for classroom 1 and classroom 2.

$$\theta \rightarrow f_X(x;\theta) \xrightarrow{x_1, x_2, \ldots, x_n} \text{Estimator} \rightarrow \hat{\theta}$$

Classical inference

- Given observations $X = (X_1, \dots, X_n)$, an estimator $\widehat{\Theta} = g(X)$ is function of X.

- Thus, $\widehat{\Theta}$ is a ____________________.

- Let n be the number of observations, the mean and variance of $\widehat{\Theta}_n$ are denoted $E_\theta\left[\widehat{\Theta}_n\right]$ and $var_\theta\left[\widehat{\Theta}_n\right]$, respectively.

Terminology regarding estimators

- Estimator: $\widehat{\Theta}_n$, a function of n observations for an $(X_1, \ldots, X_n)$ whose distribution depends on θ.
- Estimation error: $\widetilde{\Theta}_n = \widehat{\Theta}_n - \theta$.
- Bias of the estimator: $b_\theta(\widetilde{\Theta}_n) = E_\theta[\widehat{\Theta}_n] - \theta$, is the expected value of the estimation error.

Estimation of the Mean

- Suppose that the observations $X_1, \ldots, X_n$ are i.i.d., with an unknown common mean μ_X.
- $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased estimator
  - for any n, the expected value of the average is equal to the true mean.

```
heights= np.array([121.92, 132.64, 113.31, 97.20, 140.94, 139.04, 115.98, 128.27, 121.84, 97.73])
```

The average height estimate

```
average = np.mean(heights)= 122.185
```

Properties of the Estimator of the mean

- $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased estimator
  - for any n, the expected value of the average is equal to the true mean.

Estimating the variance

Let σ_X^2 denote the variance of the random variables. Then there are two cases that should be considered for estimating the variance.

Known mean: If the mean of the random variables, μ_X, is known. Let the sample-variance estimator for this case be defined by

$$\hat{\sigma_X^2} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2.$$

Estimating the variance

- Let's first determine if the sample variance estimator is biased when the true mean is known:

Estimating the variance

Known mean: If the mean of the random variables, μ_X, is known. Let the sample-variance estimator for this case be defined by

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2.$$

Unknown mean: it is natural to replace μ_X with its sample estimate $\hat{\mu}_X$:

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2.$$

Estimating the varaince

- Let's first determine if the sample variance estimator is biased when we replace the true mean with its sample estimate: (*experiment validate*)

Estimating the varaince

Known mean: If the mean of the random variables, μ_X, is known. Let the sample-variance estimator for this case be defined by

$$\hat{\sigma_X^2} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2.$$

Unknown mean: unbiased estimator:

$$\hat{\sigma_X^2} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2.$$

The change in denominator is often referred to as a *degrees-of-freedom (dof) correction*.

Example

- Suppose we have a sample of student scores from an exam, and we want to estimate the population mean score.
- Sample data: 72,85,90,88,76
- Using the same sample data, we want to estimate the population variance.

Point estimate vs interval estimate

- Instead of estimating a single value, an interval estimate is also used:

- For an unknown parameter

$$\theta \to f_X(x;\theta) \xrightarrow{x_1, x_2, \ldots, x_n} \text{Interval Estimator} \to [\hat{\theta}^-, \hat{\theta}^+]$$

An interval contain the unknown parameter with high probability

Confidence intervals (CIs)

- The value of an estimator may not be informative enough
- Let us first fix a desired confidence level, $1-\alpha$, where α is typically a small number.
- We then replace the point estimator $\widehat{\Theta}_n$ by a lower estimator $\widehat{\Theta}_n^-$ and an upper estimator $\widehat{\Theta}_n^+$, s.t.

$$P\left(\widehat{\Theta}_n^- \le \theta \le \widehat{\Theta}_n^+\right) \ge 1-\alpha$$

  for every possible value of θ.

- We call $\left[\widehat{\Theta}_n^-, \widehat{\Theta}_n^+\right]$ a $(1-\alpha)$ confidence interval.

Confidence intervals (CIs)

- $\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Recall: the observations $X_1, \ldots, X_n$ are i.i.d., with an unknown common mean μ_X

$$\hat{\mu}_X \sim \mathcal{N}(\mu_X, \frac{\sigma^2}{n})$$

- Recall CLT:

$$\text{We call } [\hat{\mu}_X^-, \hat{\mu}_X^+] \text{ a } (1-\alpha) \text{ confidence interval if}$$

$$P(\hat{\mu}_X^- \leq \mu_X \leq \hat{\mu}_X^+) > 1 - \alpha$$

Confidence intervals (CIs)

- Suppose $\alpha = 0.05$
- Let's compute the 95% confidence interval about the mean of unknown RV using the samples.

https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Confidence interval for mean estimate with unknown variance

- Recall if the variance is unknown, we have an unbiased estimate for the variance

Unknown mean: unbiased estimator:

$$\hat{\sigma_X^2} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2.$$

$$\frac{\hat{\mu}_X - \mu_X}{\sigma_X/\sqrt{n}}$$

has a Student's t-distribution with $\nu = n - 1$ degrees of freedom (dof) t_ν.
(Like a Gaussian, but more spread out!)

Summary

1. Point Estimation for Mean with prior knowledge of the population variance

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. Standard Error of the Mean with known population variance (not an estimate but can be computed based on the property of variance.)

$$SE = \frac{\sigma}{\sqrt{n}}$$

3. Point Estimation for Mean without prior knowledge of the population variance

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

4. Point Estimation for Variance without prior knowledge of the population variance

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$