ISyE 3770
Chapter 7: Statistical Estimation

Instructor: Dan Li
Slides by Prof. Nagi Gebraeel

School of Industrial and Systems Engineering
Georgia Tech

Spring 2021

1. Statistical Estimation
   - Point Estimation
   - CLT
   - Biased/Unbiased Estimators
   - MSE

Main Concepts

- Statistical inference is almost always directed to drawing conclusions about characteristics of a population. In statistical inference, a characteristic of a population is called *parameter*. Examples of parameters are the *mean* of the population or *variability* of the population, the *proportion* of a category and so on.
- Since investigators usually obtain a sample of the population, they will need to "guess" the parameters of interest based on the sample data. Evaluating the population parameters based on the sample data is called *estimation*.

Point Estimation

- A point estimate is a reasonable value of a population parameter.
- $X_1, X_2, \ldots, X_n$ are random variables.
- Functions of these random variables, $\bar{x}$ and s^2, are also random variables called statistics.
- Statistics have their unique distributions which are called sampling distributions.

Example of a Statistic

- *Sample mean* is a point estimator for the mean parameter μ:

$$\hat{\mu} = \bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

for $X_1, \ldots, X_n$ a random sample. The corresponding point estimate is:

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

for $x_1, \ldots, x_n$ observed values. Note that $\bar{x}$ is a value/number and $\bar{X}$ is a random variable.

Examples of Point Estimators

Parameter	Measure	Statistic
μ	Mean of a single population	x-bar
σ^2	Variance of a single population	s^2
σ	Standard deviation of a single population	s
p	Proportion of a single population	p-hat
$\mu_1 - \mu_2$	Difference in means of two populations	x bar$_1$ - x bar$_2$
$p_1 - p_2$	Difference in proportions of two populations	p hat$_1$ - p hat$_2$

- There could be choices for the point estimator of a parameter.
- To estimate the mean of a population, we could choose the:
  - Sample mean
  - Sample median
  - Average of the largest & smallest observations in the sample.

Point Estimation

- Conventionally, we denote the population parameters with Greek letters: μ denotes the population mean, σ^2 denotes the variability parameter and θ denotes a general parameter.
- Also we denote their point estimates or estimators by $\hat{\mu}$ for the mean, $\hat{\sigma}^2$ for the variability parameter and $\hat{\theta}$ for a general parameter.

Central Limit Theorem

- **CLT is one of the most important theorems of probability.** We will use it in various statistical inference procedures.
- **Theorem:** If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2 ($\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$), and $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$ is the average, then for large n we have the following approximations:

$$\bar{X} \approx N(\mu, \sigma^2/n)$$

  The larger the value of n is, the better the approximation.

Example

- Suppose that a random variable X has a continuous uniform distribution shown below. Find the distribution of the sample mean of a random sample of size n = 40.

$$f(n) = \begin{cases} 1/2 & 4 \leq x \leq 6 \\ 0 & \text{Otherwise} \end{cases}$$

Find the distribution of $\bar{X}$

Example

Example

- Suppose that X has a discrete uniform distribution

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3} & x = 1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

A random sample of $n = 36$ is selected from this population. Find the probability that the sample mean is greater than 2.1 but less than 2.5, assuming that the sample mean would be measured to the nearest tenth.

Example

Unbiased Estimators Defined

- We define $\hat{\theta}$ an *unbiased* estimator of θ if:

$$\mathbb{E}(\hat{\theta}) = \theta$$

- That is, the distribution of $\hat{\theta}$ is centered at θ. The difference $\mathbb{E}(\hat{\theta}) - \theta$ is called *bias* of $\hat{\theta}$.

- If two different point estimators are being compared, we will prefer the one with smaller absolute bias. So if $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators of θ then:

$$\text{Prefer } \hat{\theta}_1 \text{ if } |\mathbb{E}(\hat{\theta}_1) - \theta| < |\mathbb{E}(\hat{\theta}_2) - \theta|$$
$$\text{Prefer } \hat{\theta}_2 \text{ if } |\mathbb{E}(\hat{\theta}_1) - \theta| > |\mathbb{E}(\hat{\theta}_2) - \theta|$$

Unbiased Estimators: *Sample Mean*

- Take $X_1, \ldots, X_n$ from some distribution of mean μ and variance σ^2. That is, $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$. A point estimator of μ is:

$$\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- Is this estimator unbiased for μ?

$$\begin{aligned} \mathbb{E}[\bar{X}] &= \mathbb{E}[\frac{X_1 + X_2 + \ldots + X_n}{n}] \\ &= \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n]}{n} \\ &= \frac{[\mu + \mu + \ldots + \mu]}{n} = \mu \end{aligned}$$

Unbiased Estimators: *Sample Proportion*

- For $X \sim Bin(n, p)$, we estimate the success probability p as:

$$\hat{p} = \frac{X}{n}, \text{ given } n.$$

Is this estimator unbiased for p?

$$\mathbb{E}[\hat{p}] = \mathbb{E}[\frac{X}{n}] = \frac{1}{n}\mathbb{E}[X] = \frac{1}{n}(np) = p$$

- Therefore, no matter what the true value of p is, the distribution of the estimator $\hat{p}$ will be centered at the true value.

Unbiased Estimators: *Sample Variance*

- A point estimator of σ^2 is:

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

  A point estimator of σ is:

$$\hat{\sigma} = S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}}$$

- Is $\hat{\sigma}^2$ estimator unbiased for σ^2?

Unbiased Estimators: *Sample Variance*

Show that the sample variance S^2 is a unbiased estimator of σ^2

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}[S^2] = \mathbb{E}\left[\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}\right]$$

$$= \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}\left(X_i^2 + \bar{X}^2 - 2X_i\bar{X}\right)\right] = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2] - n\mathbb{E}\left[\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n^2}\right]\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\mathbb{V}[X_i] + \mathbb{E}[X_i]^2) - n\mathbb{E}\left[\frac{\mathbb{V}(\sum_{i=1}^{n} X_i) + (\mathbb{E}(\sum_{i=1}^{n} X_i))^2}{n^2}\right]\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - \mathbb{E}\left[\frac{n\sigma^2 + (n\mu)^2}{n}\right]\right) = \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \sigma^2$$

Unbiased Estimators Defined

- Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ. Which one to choose? Even though the distribution of each estimator is centered at the true value of θ, the spreads of the distributions about the true value θ may be different.
- Among all unbiased estimators of θ, we then choose the one that has minimum variance. The resulting estimator is called *minimum variance unbiased estimator* (MVUE) of θ.

Example

- Let X_1, X_2,..., X_7 denote a random sample from a population with mean μ and variance σ^2. Suppose that we have two estimators of μ:

$$\hat{\mu}_1 = \frac{X_1 + X_2 + ... + X_7}{7} \quad \text{and} \quad \hat{\mu}_2 = \frac{2X_1 - X_6 + X_4}{2}$$

1. Are both estimators unbiased estimators of μ?
2. What is the variance of each estimator?

Example

MSE

- What if we compare want to compare two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ such that:

$$\mathbb{E}(\hat{\theta}_1) = \mu_1,\ \mathbb{E}(\hat{\theta}_2) = \mu_2$$
$$\mathbb{V}(\hat{\theta}_1) = \sigma_1^2,\ \mathbb{V}(\hat{\theta}_2) = \sigma_2^2.$$

- We would like to compare both the bias and the variance simultaneously. To do this, we define the *mean square error*:

$$MSE(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2 = \mathbb{V}(\hat{\theta}) + (\theta - \mathbb{E}(\hat{\theta}))^2 = \mathbb{V}(\hat{\theta}) + (\text{bias})^2$$

- Therefore, we compute $MSE(\hat{\theta}_1)$ and $MSE(\hat{\theta}_2)$, and choose the estimator with the smallest mean square error.

Example

- Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of the parameter θ. We know that $\mathbb{E}(\hat{\theta}_1) = \theta$ $\mathbb{E}(\hat{\theta}_2) = \theta/2$, $\mathbb{V}(\hat{\theta}_1) = 10$, $\mathbb{V}(\hat{\theta}_2) = 4$. Which estimator is better? In what sense is it better?

Example