Chapter 7: Point Estimation of Parameters

Learning objectives

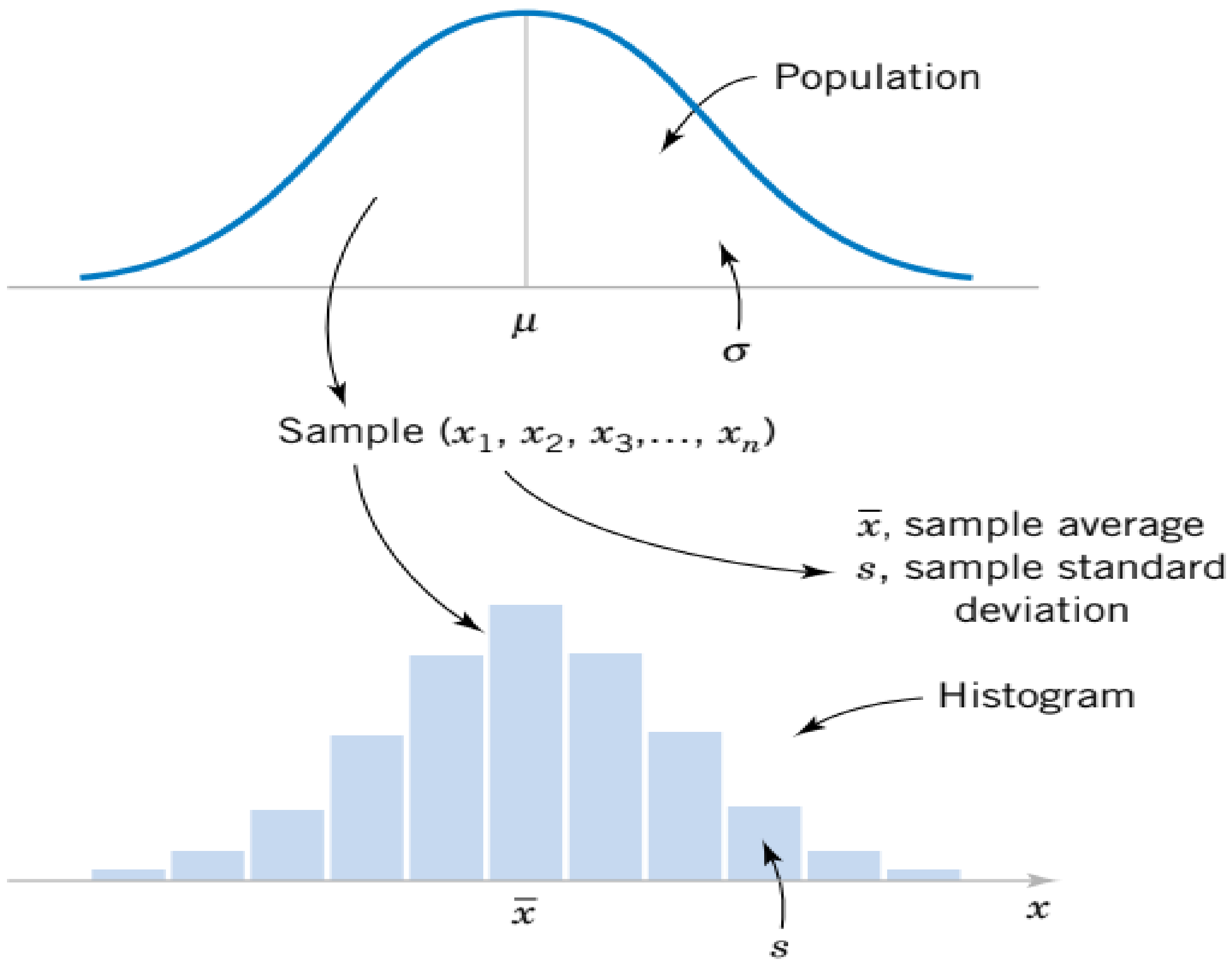
1. Introduction
2. Sampling Distributions
3. General Concepts Of Point
4. Estimation Methods Of Point Estimation

Let X is a random variable with probability distribution $f(x)$, which is characterized by the unknown $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$

For example, $X \sim N(\mu, \sigma^2)$ then $\theta = (\mu, \sigma^2)$.

How to "determine" the values of θ?

Example


Population
μ
σ
Sample ($x_1, x_2, x_3, \ldots, x_n$)
$\bar{x}$, sample average
s, sample standard deviation
Histogram
$\bar{x}$
s
x

In engineering, we often need to estimate

1. The mean μ of a single population.
2. The variance σ^2 (or standard deviation) of a single population.
3. The proportion *p* of items in a population that belong to a class of interest.
4. The difference in means of two populations, $\mu_1 - \mu_2$.
5. The difference in two population proportions, $p_1 - p_2$.

The important results on point estimation

1. For μ, the estimate is the $\hat{\mu} = \bar{x}$, sample mean.
2. For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
3. For p, the estimate is $\hat{p} = x/n$, the sample proportion.
4. For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$
5. For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$

Definition — Random Sample

The random variables $X_1, \ldots, X_n$ are called a random sample of size n if

- The X_i's are independent
- Every X_i has the same probability distribution

Definition — Statistic

- A statistic $\hat{\Theta}$ is any function of the observations $X_1, \ldots, X_n$:

$$\hat{\Theta} = h(X_1, \ldots, X_n)$$

- The probability distribution of a statistic is called a sampling distribution.

Example Two important statistic

- Sample mean $\bar{X}$

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

- Sample variance S^2

$$S^2 = \frac{(X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}$$

Theorem

In above example, if $(X_1, \ldots., X_n)$ is a random sample of size n take from a normal distribution $N(\mu, \sigma^2)$ then

- $\overline{X}$ has a normal distribution $N(\mu, \sigma^2/n)$
- $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with n-1 degrees of freedom (see pages 273-274).

Introduction

Sampling Distributions

General concepts of point estimation

Methods of point estimation
Method of moments
Method of maximum likelihood

Summary

EXAMPLE 7-1 Resistors

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

$$z = \frac{95 - 100}{2} = -2.5$$

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$

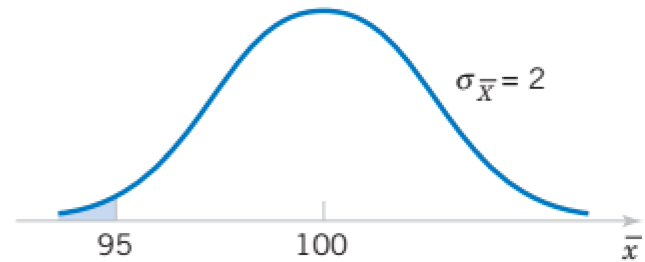


Figure 7-2 Probability for Example 7-1.

```
> #Example 7.1
> #In R, you can compute easily
> pnorm(95,100,2)
[1] 0.006209665
```

Theorem — Central Limit Theorem

- If we are sampling from a population that has an unknown probability distribution, **the sampling distribution of the sample mean $\overline{X}$ is approximately normal if the sample size is large.**
- This is one of the most useful theorems in statistics, called the **central limit theorem**.

Example: Distributions of sample means when rolling 2,3,10,…dices

Use R to simulate the rolling a dices 2,3,5,10 times and study the distribitions of sample means

- **Note that rolling a fair dice has a uniform distribution with the mean 3.5 and variance 35/12**
- **Study sampling distributions of sample** $\overline{X}$

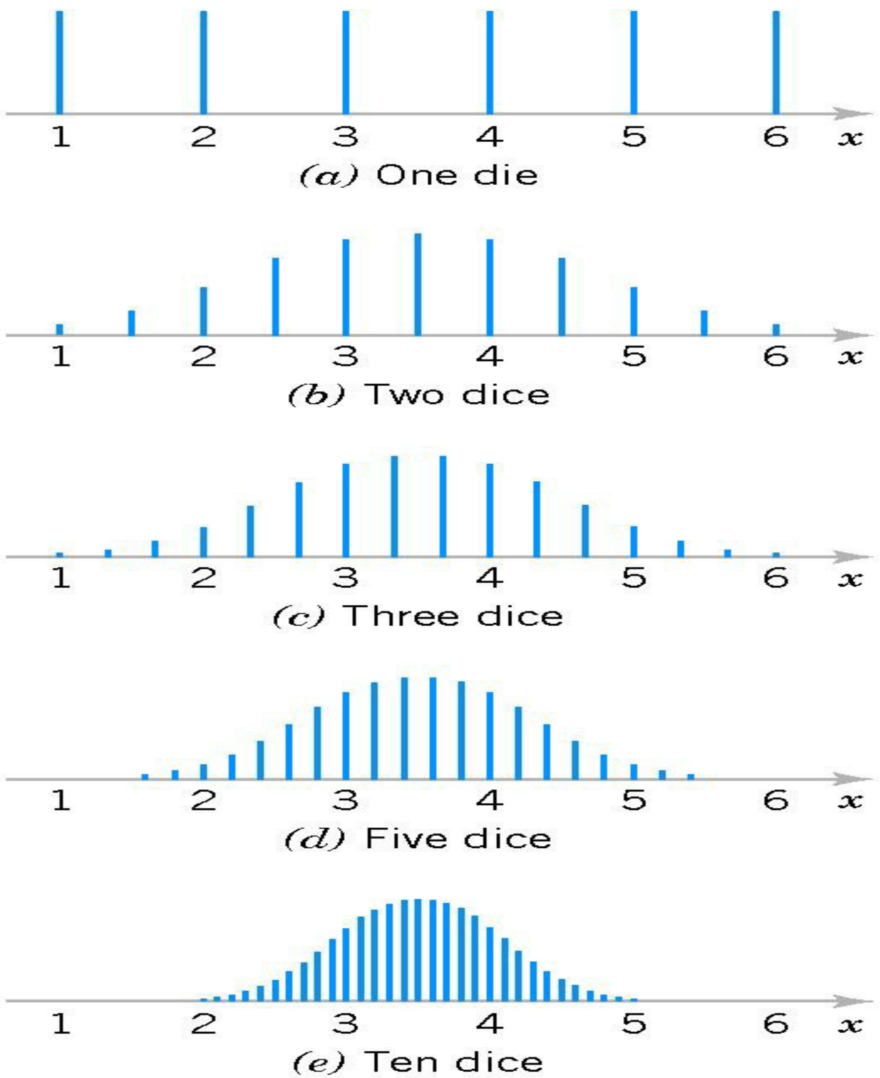
$$\mu = E(X) = (a+b)/2 \qquad \sigma^2 = \frac{(b-a+1)^2-1}{12}$$

Example: Distributions of sample means when rolling 2,3,10,…dices

Simulation with R:

1. Define a command to roll a dice n times
2. Make random samples of rolling a dice n times and compute sample means.
3. Showing the histogram of the sample means
4. Compute the mean and variance of the sample mean variable.

Figure 7-1 Distributions of average scores from throwing dice.



```r
RollDice=function(n) sample(1:6,n,replace=T)
means = vars = numeric(1000)
for (i in 1:1000) {
   samp = RollDice(2)
   #print(samp)
   means[i] = mean(samp)
   vars[i] = var(samp)
}
means
vars
hist(means)
hist(vars)
```

Theorem — **Central Limit Theorem**

Let $(X_1, \ldots., X_n)$ is a random sample of size n take from a population with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution.

Remark: The normal approximation for $\overline{X}$ depends on the sample size n.

EXAMPLE 7-2 Central Limit Theorem

Suppose that a random variable X has a continuous uniform distribution

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.

The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6-4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of $\overline{X}$ is approximately normal with mean $\mu_{\overline{X}} = 5$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n = 1/[3(40)] = 1/120$. The

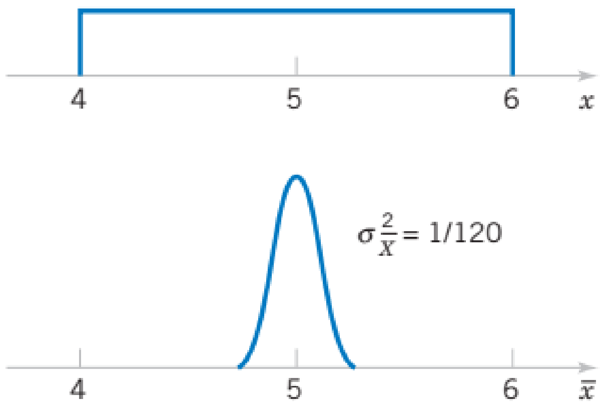


Figure 7-3 The distributions of X and $\overline{X}$ for Example 7-2.

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{7-4}$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

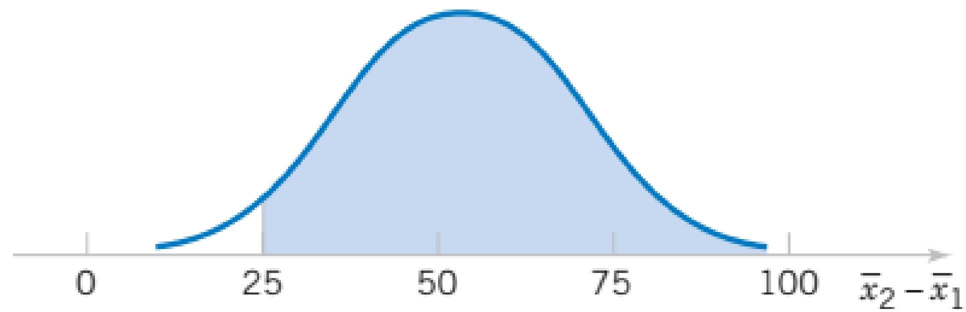
Approximate Sampling Distribution of a Difference in Sample Means

EXAMPLE 7-3 Aircraft Engine Life

The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. What is the probability that the difference in the two sample means $\overline{X}_2 - \overline{X}_1$ is at least 25 hours? Assume that the old and improved processes can be regarded as independent populations.

Approximate Sampling Distribution of a Difference in Sample Means

Figure 7-4 The sampling distribution of $\overline{X}_2 - \overline{X}_1$ in Example 7-3.



To solve this problem, we first note that the distribution of $\overline{X}_1$ is normal with mean $\mu_1 = 5000$ hours and standard deviation $\sigma_1/\sqrt{n_1} = 40/\sqrt{16} = 10$ hours, and the distribution of $\overline{X}_2$ is normal with mean $\mu_2 = 5050$ hours and standard deviation $\sigma_2/\sqrt{n_2} = 30/\sqrt{25} = 6$ hours. Now the distribution of $\overline{X}_2 - \overline{X}_1$ is normal with mean $\mu_2 - \mu_1 = 5050 - 5000 = 50$ hours and variance $\sigma_2^2/n_2 + \sigma_1^2/n_1 = (6)^2 + (10)^2 = 136$ hours^2. This sampling distribution is shown in Fig. 7-4. The probability that $\overline{X}_2 - \overline{X}_1 \geq 25$ is the shaded portion of the normal distribution in this figure.

Approximate Sampling Distribution of a Difference in Sample Means

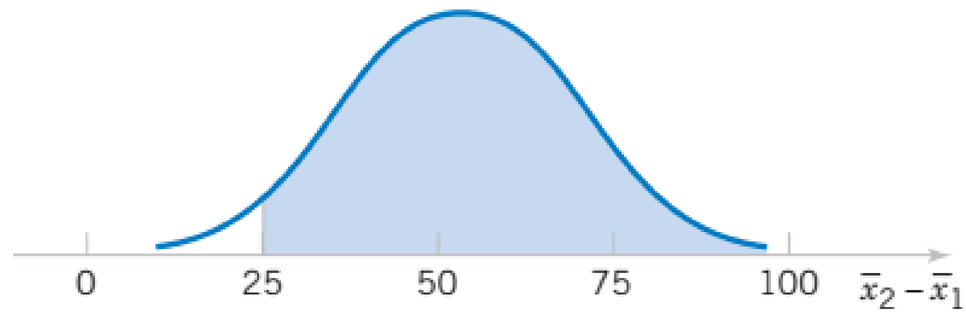


Figure 7-4 The sampling distribution of $\bar{X}_2 - \bar{X}_1$ in Example 7-3.

Corresponding to the value $\bar{x}_2 - \bar{x}_1 = 25$ in Fig. 7-4, we find that

$$z = \frac{25 - 50}{\sqrt{136}} = -2.14$$

and consequently,

$$P(\bar{X}_2 - \bar{X}_1 \geq 25) = P(Z \geq -2.14)$$
$$= 0.9838$$

Therefore, there is a high probability (0.9838) that the difference in sample means between the new and the old process will be at least 25 hours if the sample sizes are $n_1 = 16$ and $n_2 = 25$.

Definition Point Estimate

- A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$.
- The statistic $\hat{\Theta}$ is called the **point estimator**.

Two steps to find point estimation:

Step 1. Determine $\hat{\Theta}$ by using the theoretical results.

Steps 2. Calculate $\hat{\theta}$ from the experimental data.

Example

Market researchers use the number of sentences per advertisement as a measure of readability for magazine advertisements. The following represents a random sample of the number of sentences found in 15 advertisements. Find a point estimate of the population mean, μ.

9 20 18 16 9 9 11 13 22 16 5 18 6 6 5

Step 1. $$\hat{\Theta} = \frac{X_1 + \ldots + X_{15}}{15}$$

Step 2. A point estimate for μ is

$$\hat{\mu} = \frac{9+20+\ldots+6+5}{15} = \frac{183}{15} = 12.2$$

Definition — Unbiased Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

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

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta} - \theta)$$

is called the **bias** of the estimator $\hat{\Theta}$.

Example

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\bar{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

For the sample mean.

$$E(\bar{X}) = E\frac{X_1 + \ldots + X_n}{n} = \frac{EX_1 + \ldots + EX_n}{n} = \mu$$

Now consider the sample variance. We have

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

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

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

Hence $E(S^2)=\sigma^2$

Definition Minimum Variance Unbiased Estimator

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

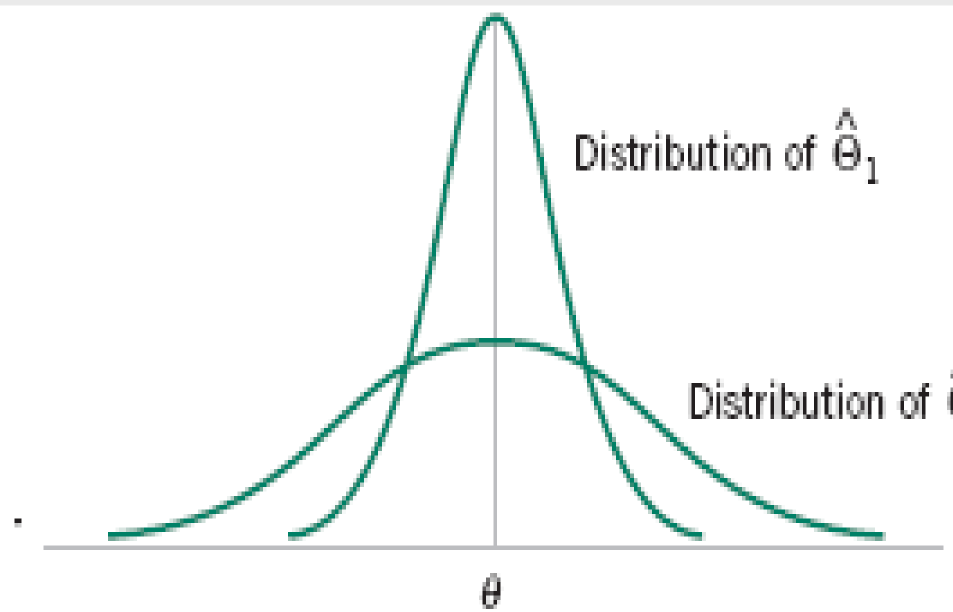


Figure 7-5
The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

Theorem

Let $(X_1, \ldots, X_n)$ is a random sample of size n take from a normal distribution $N(\mu, \sigma^2)$, the sample mean $\overline{X}$ is the MVUE for μ.

Definition Minimum Variance Unbiased Estimator

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by

$$\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$$

If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\hat{\Theta}}$

Remark

If the random sample of size n take from a normal distribution $N(\mu, \sigma^2)$ then $\overline{X}$ has a normal distribution $N(\mu, \sigma^2/n)$.

Standard error of $\overline{X}$ is

$$\sigma_{\hat{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ then the estimated standard error of $\overline{X}$ would be

$$\hat{\sigma}_{\hat{X}} = \frac{s}{\sqrt{n}}$$

Exercise

A new method of measuring the thermal conductivity of Armco iron: Using a temperature of 100^0F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-^{0}F) were obtained:

41.60, 41.48, 42.34, 41.95, 41.86,

42.18, 41.72, 42.26, 41.81, 42.04

a, Find a point estimate of the mean thermal conductivity at 100^0F and 550 watts.

b, Find the estimated standard error of $\overline{X}$

Definition — The kth moments

Let $(X_1, \ldots, X_n)$ be a random sample from random variable X with the probability distribution $f(x)$

- The kth moment of X are denoted by $M_k := E(X^k)$
- The corresponding kth **sample moment** is

$$\overline{X}^k = \frac{X_1^k + \ldots + X_n^k}{n}$$

Definition — The kth moments

Let $(X_1, \ldots, X_n)$ be a random sample from either a probability mass function or probability density function with m unknown parameters $(\theta_1, \theta_2, \ldots, \theta_m)$. The moment estimators $\hat{\Theta}_1, \ldots, \hat{\Theta}_m$ are found the following system

$$\begin{cases} M_1 = \overline{X}^1 \\ \ldots\ldots\ldots\ \ldots \\ M_m = \overline{X}^m \end{cases}$$

Example

Suppose that $X_1, \ldots, X_n$ be a random sample of size n from a normal distribution $N(\mu, \sigma^2)$. Find the moment estimators of μ and σ^2.

We have to solve following system

$$\begin{cases} M_1 = \overline{X}^1 \\ M_2 = \overline{X}^2 \end{cases} \quad \text{or} \quad \begin{cases} EX = \dfrac{X_1 + \ldots + X_n}{n} \\ EX^2 = \dfrac{X_1^2 + \ldots + X_n^2}{n} \end{cases}$$

Here $E(X) = \mu$ and $E(X^2) = V(X) + (EX)^2 = \sigma^2 + \mu^2$

Hence, by solving the above system we have

$$\mu = \overline{X}, \qquad \sigma^2 = \frac{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Finally, the moment estimators are

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Unbiased estimator

Biased estimator

Definition Maximum likelihood Estimator

Suppose that X is a random variable with probability distribution $f(x, \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, , x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1, \theta)\, f(x_2, \theta) \ldots f(x_n, \theta)$$

The maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the $L(\theta)$.

Example

Let X is a Bernoulli random variable. The probability mass function is

$$f(x,p)=\begin{cases} p^x(1-p)^{1-x}, & x=0,1 \\ 0 & otherwise \end{cases}$$

Find the maximum likelihood estimator of p.

The likelihood function of a random sample of size n is

$$L(p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

The maximum likelihood estimator of p is

$$\hat{P} = \frac{X_1 + \ldots + X_n}{n}$$

Is this a unbiased estimator?

Exercise

Let X is a random variable, which has probability mass function is

$$f(x)=\begin{cases} p & , \quad x=-1 \\ 1-p & , \quad x=1 \\ 0 & , \quad otherwise \end{cases}$$

Find the maximum likelihood estimator of *p*.

Solution:

$$f(x)=\begin{cases} p^{\frac{1-x}{2}}(1-p)^{\frac{1+x}{2}} & , \quad x=-1,1 \\ 0 & , \quad otherwise \end{cases}$$

$$\hat{P}=\frac{1}{2}-\frac{X_1+\ldots+X_n}{2n}$$

Invariance Property

Let $\hat{\Theta}_1, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \ldots, \hat{\Theta}_k$.

Exercise

Let X be normally distributed $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown.

(a) Find the maximum likelihood estimator of μ and σ^2.

(b) Find the maximum likelihood estimator of $\theta = \mu^2\sigma^2$.

Solution:

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

$$\hat{\theta} = (\overline{X})^2 \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

SUMMARY

We have studied:

1. Sampling distributions
2. Central Limit Theorem
3. General concepts of point estimation: Unbiased Estimator
4. Methods of point estimation

- Method of Moments
- Method of Maximum Likelihood

Homework: Read slides of the next lecture.