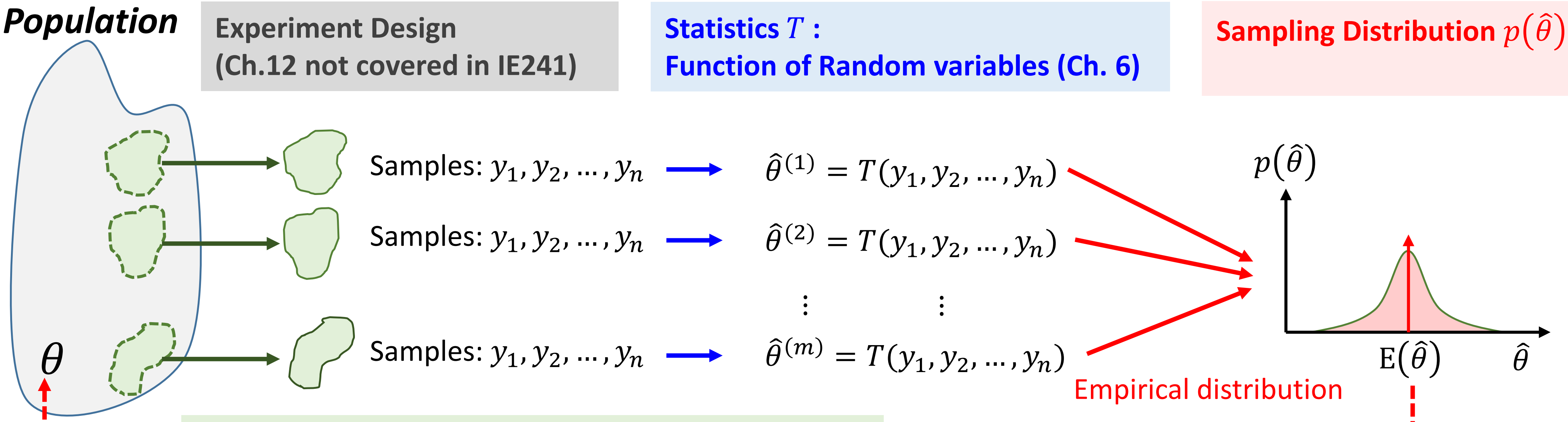
CHAPTER 8
ESTIMATION

Road Map on IE241 I

Experiment Design (Ch.12 not covered in IE241)

Statistics T : Function of Random variables (Ch. 6)

Sampling Distribution $p(\hat{\theta})$: (Ch. 7)



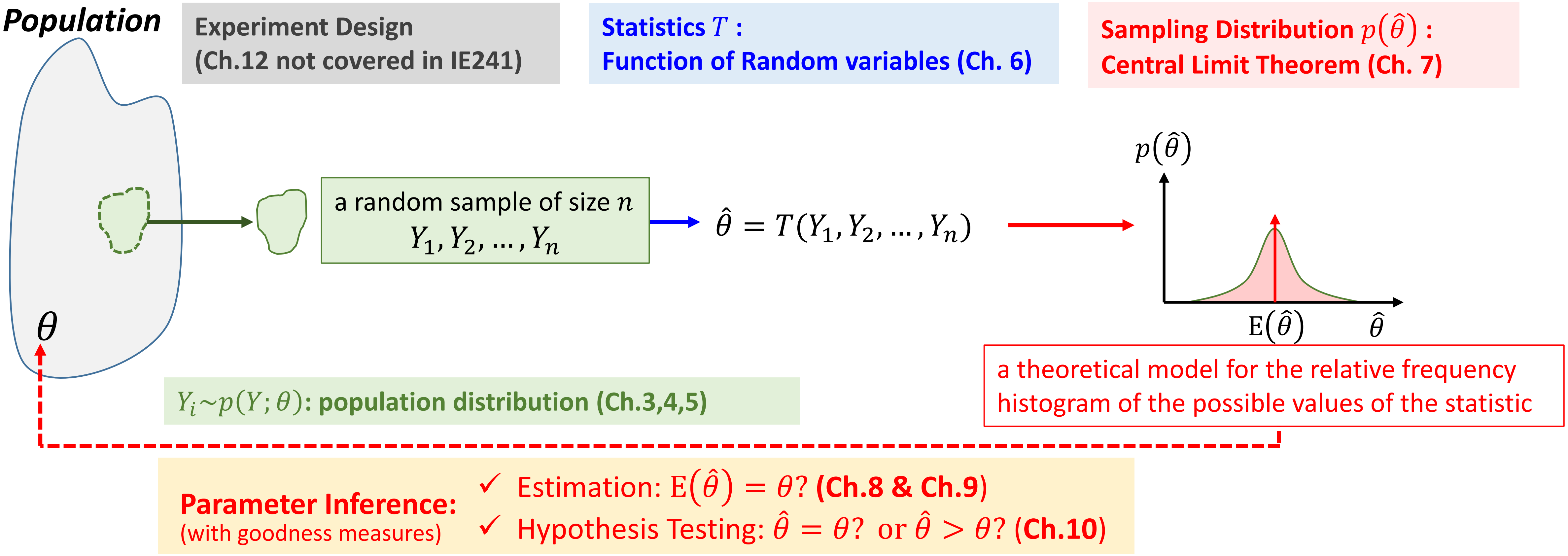
$Y_i \sim p(Y;\theta)$: **population distribution (Ch.3,4,5)**

Parameter Inference:
(with goodness measures)

- ✓ Estimation: $\mathrm{E}(\hat{\theta}) = \theta$? **(Ch.8 & Ch.9)**
- ✓ Hypothesis Testing: $\hat{\theta} = \theta$? or $\hat{\theta} > \theta$? **(Ch.10)**

- **Probability Theory (Ch.2 ~ Ch.6) plays an important role in inference** by computing the probability of the occurrence of the sample and connects the computed probability to the most probable target parameter.
- ***Estimator*** $\hat{\theta} = T(Y_1, Y_2, \dots, Y_n)$ for a target parameter θ is a function of the random variables observed in a sample and therefore itself is a random variable.
- **Sampling distribution** $p(\hat{\theta})$ can be used to evaluate the goodness of the ***estimator*** (confidence interval) and the errors (i.e., α and β errors) of ***hypothesis testing***.

Road Map on IE241 I



- **Probability Theory (Ch.2 $\sim$ Ch.6) plays an important role in inference** by computing the probability of the occurrence of the sample and connects the computed probability to the most probable target parameter.
- ***Estimator*** $\hat{\theta} = T(Y_1, Y_2, \dots, Y_n)$ for a target parameter θ is a function of the random variables observed in a sample and therefore itself is a random variable.
- **Sampling distribution** $p(\hat{\theta})$ can be used to evaluate the goodness of the ***estimator*** (confidence interval) and the errors (i.e., α and β errors) of ***hypothesis testing***.

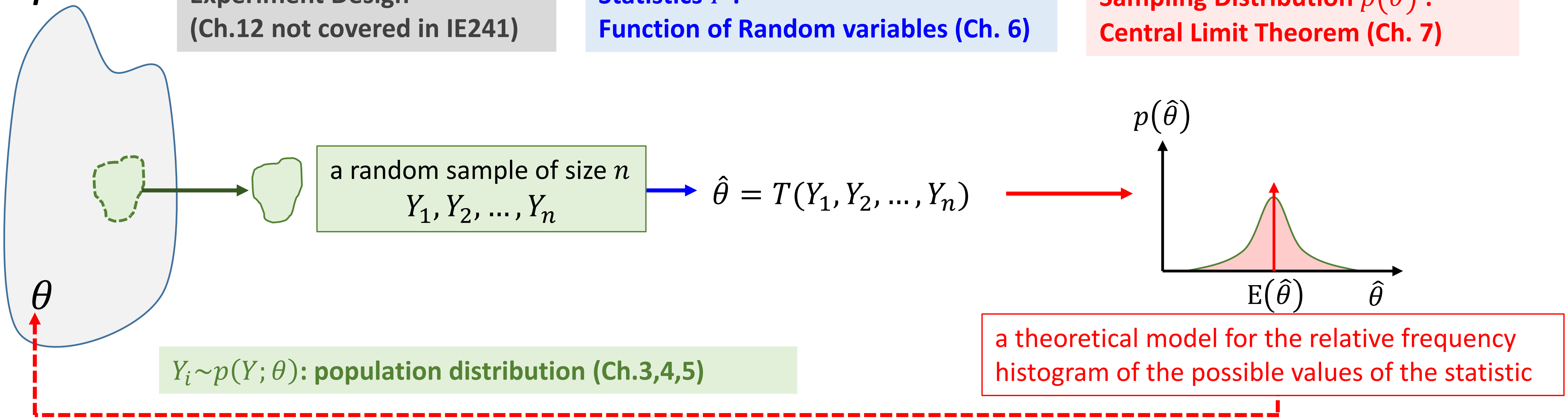
Overview on Estimation

Population

Experiment Design
(Ch.12 not covered in IE241)

Statistics T :
Function of Random variables (Ch. 6)

Sampling Distribution $p(\hat{\theta})$:
Central Limit Theorem (Ch. 7)

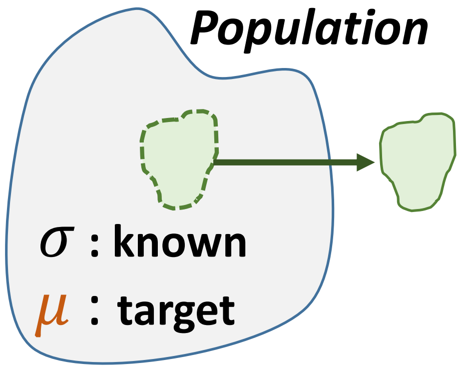


- An estimator $\hat{\theta} = E(Y_1, Y_2, \dots, Y_n)$ is a Statistics $T(Y_1, Y_2, \dots, Y_n)$
- Type of estimator $\hat{\theta} = E(Y_1, Y_2, \dots, Y_n)$
  - Point estimator vs. Interval estimator
- Goodness of estimator
  - Bias vs. variance
- Impact of the size of sample
  - Large sample size vs. Small sample size

Motivation

- An estimator $\hat{\theta} = E(Y_1, Y_2, \ldots, Y_n)$ is a Statistics $T(Y_1, Y_2, \ldots, Y_n)$
- Type of estimator $\hat{\theta} = E(Y_1, Y_2, \ldots, Y_n)$
  - Point estimator vs. Interval estimator
- Goodness of estimator
  - Bias vs. variance
- Impact of the size of sample
  - Large sample size vs. Small sample size

Estimation of the mean μ for the population distribution

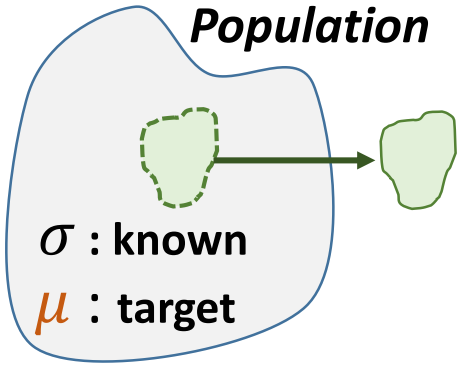


a random sample of size n: $Y_1, Y_2, \dots, Y_n$ → **Sample mean** $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ → $\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$

- When $Y_i \sim N(\mu, \sigma^2)$
- When n is **small or large**
- When σ is assumed to be **known**

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$$

Estimation of the mean μ for the population distribution



a random sample of size n
$Y_1, Y_2, \ldots, Y_n$

Sample mean

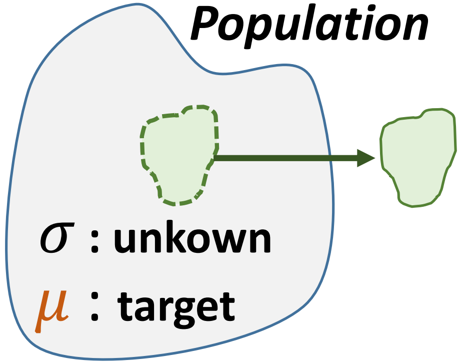
$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$$

- When $Y_i \sim$ Distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$
- When n is **large**
- When σ is assumed to be **known**

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) \sim Z \sim N(0, 1^2)$$ Due to Central limit theorem

Estimation of the mean μ for the population distribution



a random sample of size n

$Y_1, Y_2, \dots, Y_n$

Sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

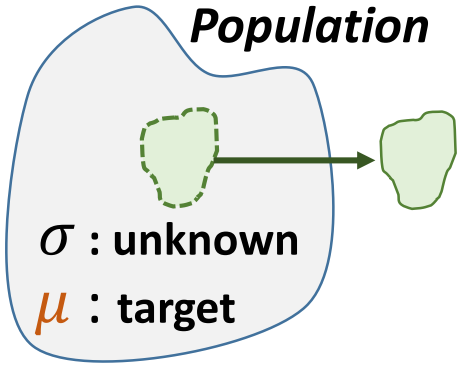
Sample variance

- When $Y_i \sim$ Distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$
- When n is **large**
- When σ is assumed to be **unknown** (s is used instead of σ)

$$\left(\frac{\bar{Y}-\mu}{\sigma_{\bar{Y}}}\right) = \left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}\right) \approx \left(\frac{\bar{Y}-\mu}{s/\sqrt{n}}\right) \sim Z \sim N(0, 1^2)$$

Due to Central limit theorem

Estimation of the mean μ for the population distribution



a random sample of size n
$Y_1, Y_2, \dots, Y_n$

Sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Sample variance

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = Z \sim N(0, 1^2)$$

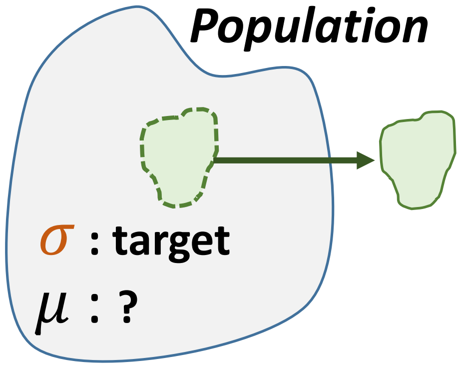
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2, df = (n-1)$$

$$\frac{Z}{\sqrt{W/v}} = \sqrt{n}\left(\frac{\bar{Y}-\mu}{S}\right) \sim t, df = (n-1)$$

- When $Y_i \sim N(\mu, \sigma^2)$
- When n is **small**
- When σ is assumed to be **unknown**

$$\sqrt{n}\left(\frac{\bar{Y}-\mu}{S}\right) \sim t, df = (n-1)$$

Estimation of the variance σ^2 for the population distribution



a random sample of size n
$Y_1, Y_2, \ldots, Y_n$

Sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2, df = (n-1)$$

- When $Y_i \sim N(\mu, \sigma^2)$
- When n is **small or large**

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2, df = (n-1)$$

Practical Uses of Estimation

- Estimation has many practical applications. For example,
  - ✓ estimating the proportion p of washers that can be expected to fail prior to the expiration of a 1-year guarantee time.
  - ✓ Other important population parameters are the population mean, variance, and standard deviation. For example,
    - ➢ the mean waiting time μ at a supermarket checkout station or
    - ➢ the standard deviation of the error of measurement σ of an electronic instrument.

Types of an Estimation

- Suppose that we wish to estimate the average amount of mercury μ that a newly developed process can remove from 1 ounce of ore obtained at a geographic location

  ✓ A point estimate : 13 ounce-that we think is close to the unknown population mean μ

  ✓ An interval estimate : (.07, .19) that is intended to enclose the parameter of interests

- The information in the sample can be used to calculate the value of a point estimate, an interval estimate, or both.

- In any case, the actual estimation is accomplished by using an estimator for the target parameter.

Definition of Estimator

DEFINITION 8.1

An *estimator* is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.

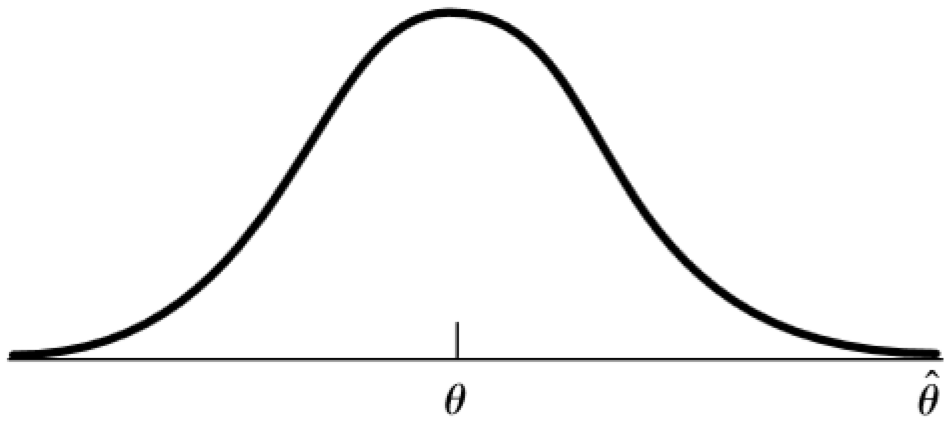
- Many different estimators (rules for estimating) may be obtained for the same population parameter.
  - ✓ Ten engineers, each assigned to estimate the cost of a large construction job, could use different methods of estimation and thereby arrive at different estimates of the total cost.
    - ➢ Each estimator represents a unique human subjective rule for obtaining a single estimate.
    - ➢ The management of a construction firm must define good and bad as they relate to the estimation of the cost of a job.
- How can we establish criteria of **goodness** to compare statistical estimators

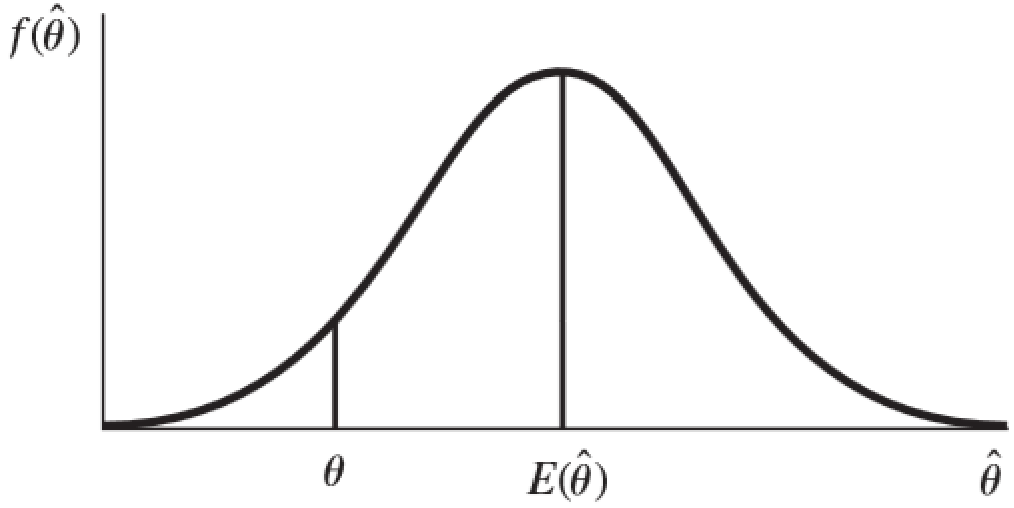
Motivation

- Suppose that a man fires **a single shot** at a target and that shot pierces the bull'seye.
  - we would not decide that the man is an expert Marksperson based on such a small amount of evidence.
- On the other hand, if **100 shots** in succession hit the bull's-eye,
  - we might acquire sufficient confidence in the marksperson and consider holding the target for the next shot if the compensation was adequate.
- The point is that we cannot evaluate the goodness of a point estimation procedure on the basis of the value of a single estimate;
  - rather, we must observe the results when the estimation procedure is used many, many times.

Goodness of an Estimator

- Because the estimates are numbers, we evaluate **the goodness of the point estimator** by constructing a frequency distribution of the values of the estimates obtained **in repeated sampling** and note how closely this distribution clusters about the target parameter.
- Suppose that we wish to specify a point estimate for a population parameter that we will call θ.
  - ✓ The estimator of $\bar{\theta}$ will be indicated by the symbol $\bar{\theta}$, read as "θ hat."





Which one is a good estimator?

Unbiased Estimator

DEFINITION 8.2

Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an *unbiased estimator* if

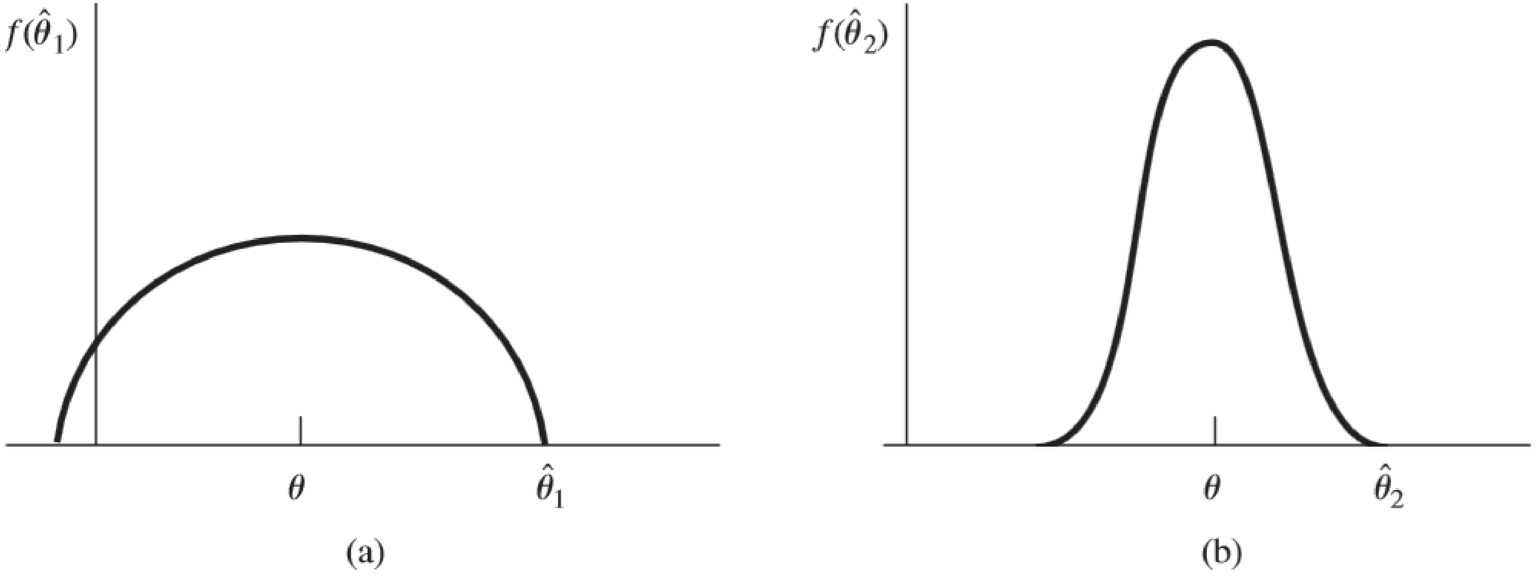
$$E(\hat{\theta}) = \theta.$$

If $E(\hat{\theta}) \neq \theta$, $\hat{\theta}$ is said to be *biased*.

DEFINITION 8.3

The *bias* of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Variance of Estimates



Which one is a good estimator?

- Figure shows two possible sampling distributions for unbiased point estimators for a target parameter θ.
- We would prefer that our estimator have the type of distribution indicated in (b) because
  - ✓ the smaller variance guarantees that in repeated sampling a higher fraction of values of $\bar{\theta}_2$ will be “close” to θ.
- Thus, in addition to preferring unbiasedness, we want the variance of the distribution of the estimator $V(\bar{\theta})$ to be as small as possible.

Motivation

DEFINITION 8.4

The *mean square error* of a point estimator $\hat{\theta}$ is

$$\mathrm{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right].$$

- It can be shown that

$$\mathrm{MSE}(\hat{\theta}) = V(\hat{\theta}) + \left[B(\hat{\theta})\right]^2$$

Proof:

Motivation

- It seems natural
  - to use the sample mean $\bar{Y}$ to estimate the population mean μ and
  - to use the sample proportion $\hat{p} = Y/n$ to estimate a binomial parameter p.

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \longrightarrow \mu$$

$$\bar{p} = \frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow p$$

Motivation

- If an inference is to be based on independent random samples of n_1 and n_2 observations selected from two different populations,
  - to use the difference between sample means $\bar{Y}_1 - \bar{Y}_2$ to estimate the difference between the population mean $\bar{\mu}_1 = \bar{\mu}_2$
  - to use the difference between sample means $\bar{p}_1 - \bar{p}_2$ to estimate the difference between the two binomial parameters $p_1 - p_2$

$$\bar{Y}_1 - \bar{Y}_2 \longrightarrow \mu_1 - \mu_2$$

$$\hat{p}_1 - \hat{p}_2 \longrightarrow p_1 - p_2$$

Contents

when random sampling has been employed (each sample is independent to each other)

$$E(\bar{Y}_1 - \bar{Y}_2) = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2$$

$$V(\bar{Y}_1 - \bar{Y}_2) = V(\bar{Y}_1) + V(\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$$

$$V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Notations:

$\sigma_{\hat{\theta}}^2$: the variance of the sampling distribution of the estimator $\hat{\theta}$.

$\sigma_{\hat{\theta}} = \sqrt{\sigma_{\hat{\theta}}^2}$: the standard deviation of the sampling distribution of the estimator $\hat{\theta}$.

usually called **the standard error** of the estimator $\hat{\theta}$.

Common Point Estimates

Table 8.1 Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\dfrac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \dfrac{Y}{n}$	p	$\sqrt{\dfrac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}^{*\dagger}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}^{\dagger}$

$^*\sigma_1^2$ and σ_2^2 are the variances of populations 1 and 2, respectively.

†The two samples are assumed to be independent.

- First, the expected values and standard errors for Y and Y1 −Y2 given in the table are valid regardless of the distribution of the population(s) from which the sample(s) is (are) taken.
- Second, all four estimators possess probability distributions that are approximately normal for large samples.

Example

EXAMPLE 8.1

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$.

Show that

$$S'^2 = \frac{1}{n}\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2$$

is a **biased estimator** for σ^2 and that

$$S^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2$$

is an **unbiased estimator** for σ^2.

Example

SOLUTION 8.1

$$\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2 = \Sigma_{i=1}^{n} Y_i^2 - \frac{1}{n}(\Sigma_{i=1}^{n} Y_i)^2 = \Sigma_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

Thus,

$$E[\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2] = E\left[\Sigma_{i=1}^{n} Y_i^2\right] - nE[\bar{Y}^2] = \Sigma_{i=1}^{n} E\left[Y_i^2\right] - nE[\bar{Y}^2] = (n-1)\sigma^2$$

$$\left(\because\ E\left(Y_i^2\right) = V(Y_i) + E(Y_i)^2 = \sigma^2 + \mu^2, \qquad E[\bar{Y}^2] = V(\bar{Y}) + E(\bar{Y})^2 = \frac{\sigma^2}{n} + \mu^2\right)$$

It follows that

$$E(S'^2) = \frac{1}{n} E[\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2] = \frac{1}{n}(n-1)\sigma^2 = \frac{n-1}{n}\sigma^2\ (biased)$$

$$E(S^2) = \frac{1}{n-1} E[\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2\ (unbiased)$$

Motivation

If we use an estimator once and acquire a single estimate, how good will this estimate be?
How much faith can we place in the validity of our inference

- One way to measure the goodness of any point estimation procedure is in terms of the distances between the estimates that it generates and the target parameter.

- This quantity, which varies randomly in repeated sampling, is called the error of estimation. Naturally we would like the error of estimation to be as small as possible.

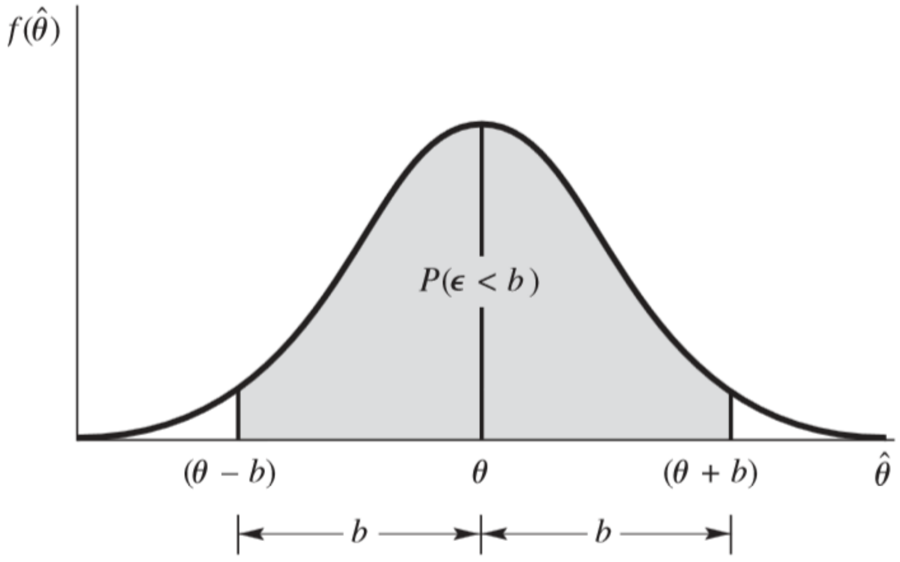
Definition

DEFINITION 8.5

The *error of estimation* ε is the distance between an estimator and its target parameter. That is, $\varepsilon = |\hat{\theta} - \theta|$.

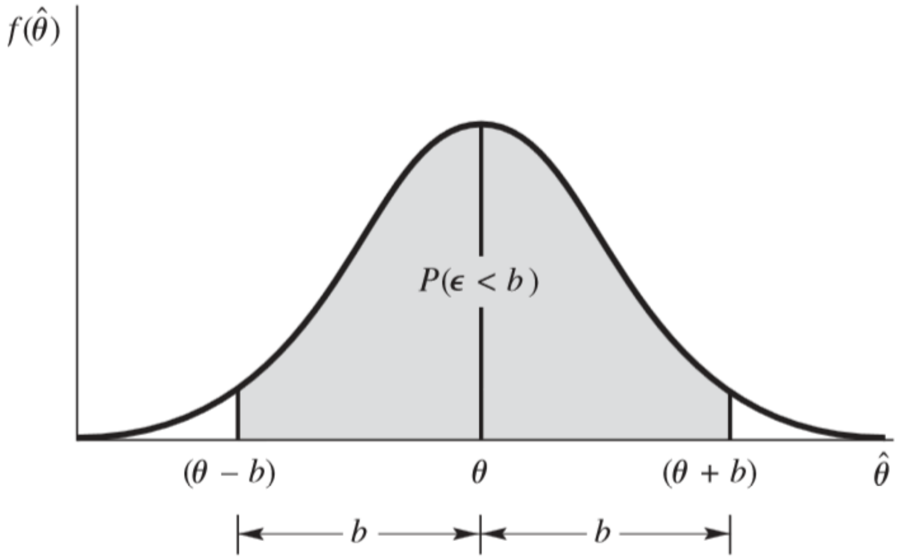
- Because $\hat{\theta}$ is a random variable, the error of estimation is also a random quantity, and
- We cannot say how large or small it will be for a particular estimate. However, we can make probability statements about it
  - ✓ $P\left(\left|\hat{\theta} - \theta\right| < b\right) = P[-b < \bar{\theta} - \theta < b] = P(\theta - b < \bar{\theta} < \theta + b)$
    - ➢ We can think of b as a probabilistic bound on the error of estimation.

Bound on Point Estimate



- Although we are not certain that a given error is less than b (because ε is R.V.), the above figure indicates that $P(\varepsilon < b)$ is high.
  - If b can be regarded from a practical point of view as small, then $P(\varepsilon < b)$ provides a measure of the goodness of a single estimate.
  - This probability identifies the fraction of times, in repeated sampling, that the estimator $\bar{\theta}$ falls within b units of θ, the target parameter.

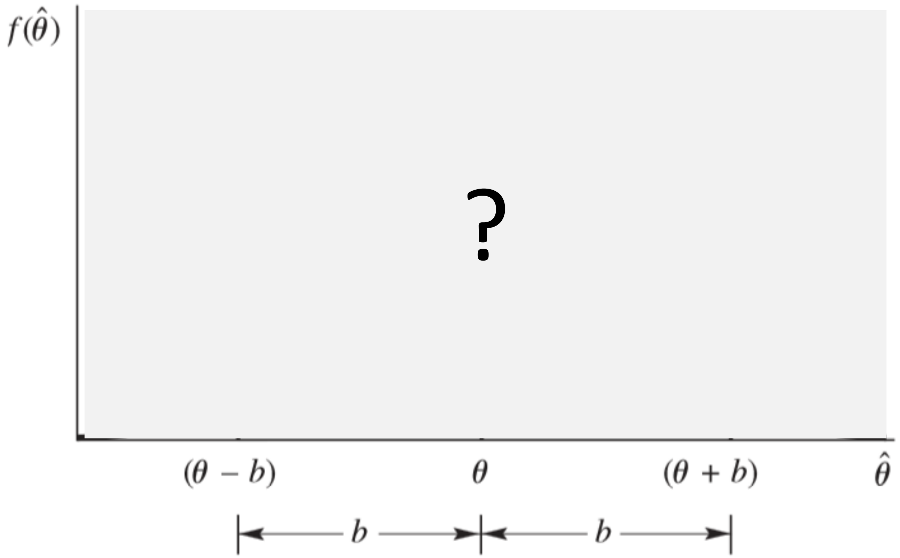
Bound on Point Estimate



- Suppose that we want to find the value of b so that $P(\varepsilon < b) = .90$.
- If we know the probability density function of $f(\bar{\theta})$ on $\bar{\theta}$, then we can find b such that

$$P(\varepsilon < b) = \int_{\theta-b}^{\theta+b} f(\bar{\theta})d\bar{\theta} = 0.90$$

Bound on Point Estimate



- Suppose that we want to find the value of b so that $P(\varepsilon < b) = .90$.
- But whether we know the probability distribution of $\bar{\theta}$ or not, if $\bar{\theta}$ is unbiased we can find an approximate bound on ε by expressing b as a multiple of the standard error of $\bar{\theta}$

$$P(\epsilon < b) = P(\epsilon < k\sigma_{\bar{\theta}}) = P\left(\left|\bar{\theta} - \theta\right| < k\sigma_{\bar{\theta}}\right) \geq 1 - \frac{1}{k^2}$$

due to Tchebysheff's Theorem

Recall

THEOREM 4.13 (Tchebysheff's Theorem)

Let Y be a random variable with finite mean μ and variance σ^2. Then, for any $k > 0$,

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad or \quad P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Bound on Point Estimate

Table 8.2 Probability that $(\mu - 2\sigma) < Y < (\mu + 2\sigma)$

Distribution	Probability
Normal	.9544
Uniform	1.0000
Exponential	.9502

- The point is that $b = 2\sigma_{\hat{\theta}}$ is a good approximate bound on the error of estimation in most practical situations.
- According to Tchebysheff's theorem, the probability that the error of estimation will be less than this bound $(\theta \pm 2\sigma_{\hat{\theta}})$ is at least .75.
  - ✓ the bounds for probabilities provided by Tchebysheff's theorem are usually very conservative;
  - ✓ the actual probabilities usually **exceed** the Tchebysheff bounds by a considerable amount.

Example

EXAMPLE 8.2

A sample of $n = 1000$ voters, randomly selected from a city, showed $y = 560$ in favor of candidate Jones. Estimate p, the fraction of voters in the population favoring Jones, and place a 2-standard-error bound on the error of estimation.

Example

SOLUTION 8.2

$$\hat{p} = \frac{y}{n} = \frac{560}{1000} = 0.56$$

The probability distribution of $\hat{p}$ is very accurately approximated by a normal probability distribution for large samples. Since $n = 1000$, when $b = 2\sigma_{\hat{p}}$, the probability that ε will be less than b is approximately 0.95. From Table 8.1, the standard error of the estimator for p is given by $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$. Therefore, $b = 2\sigma_{\hat{p}} = 2\sqrt{pq/n}$.

Unfortunately, to calculate *b*, we need to know *p*, and estimating *p* was the objective of our sampling. This apparent stalemate is not a handicap, however, because $\sigma_{\hat{p}}$ varies little for small changes in p. Hence, substitution of the estimate $\hat{p}$ for p produces little error in calculating the exact value of $b = 2\sigma_{\hat{p}}$. Then, for our example, we have

$$b = 2\sigma_{\hat{p}} = 2\sqrt{pq/n} \approx 2\sqrt{\frac{(0.56)(0.44)}{1000}} = 0.03$$

What is the significance of our calculations? The probability that the error of estimation is less than 0.03 is approximately 0.95. Consequently, we can be reasonably confident that our estimate, 0.56, is within 0.03 of the true value of *p*, the proportion of voters in the population who favor Jones.

Example

EXAMPLE 8.3

A comparison of the durability of two types of automobile tires was obtained by road testing samples of $n_1 = n_2 = 100$ tires of each type. The number of miles until wear-out was recorded, where wear-out was defined as the number of miles until the amount of remaining tread reached a prespecified small value. The measurements for the two types of tires were obtained independently, and the following means and variances were computed:

$$\bar{y}_1 = 26{,}400\ miles, \qquad \bar{y}_2 = 25{,}100\ miles,$$
$$s_1^2 = 1{,}440{,}000, \qquad s_2^2 = 1{,}960{,}000.$$

Estimate the difference in mean miles to wear-out and place a 2-standard-error bound on the error of estimation.

Example

SOLUTION 8.3

The point estimate of $(\mu_1 - \mu_2)$ is $(\bar{y}_1 - \bar{y}_2) = 26{,}400 - 25{,}100 = 1300\ miles,$

and the standard error of the estimator (see Table 8.1) is $\sigma_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$

We must know σ_1^2 and σ_2^2, or have good approximate values for them, to calculate $\sigma_{\bar{Y}_1-\bar{Y}_2}$. Fairly accurate values of σ_1^2 and σ_2^2 often can be calculated from similar experimental data collected at some prior time, or they can be obtained from the current sample data by using the unbiased estimators

$$\hat{\sigma}_i^2 = S_i^2 = \frac{1}{n_i - 1}\Sigma_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_i\right)^2, \qquad i = 1,2$$

These estimates will be adequate if the sample sizes are reasonably large.

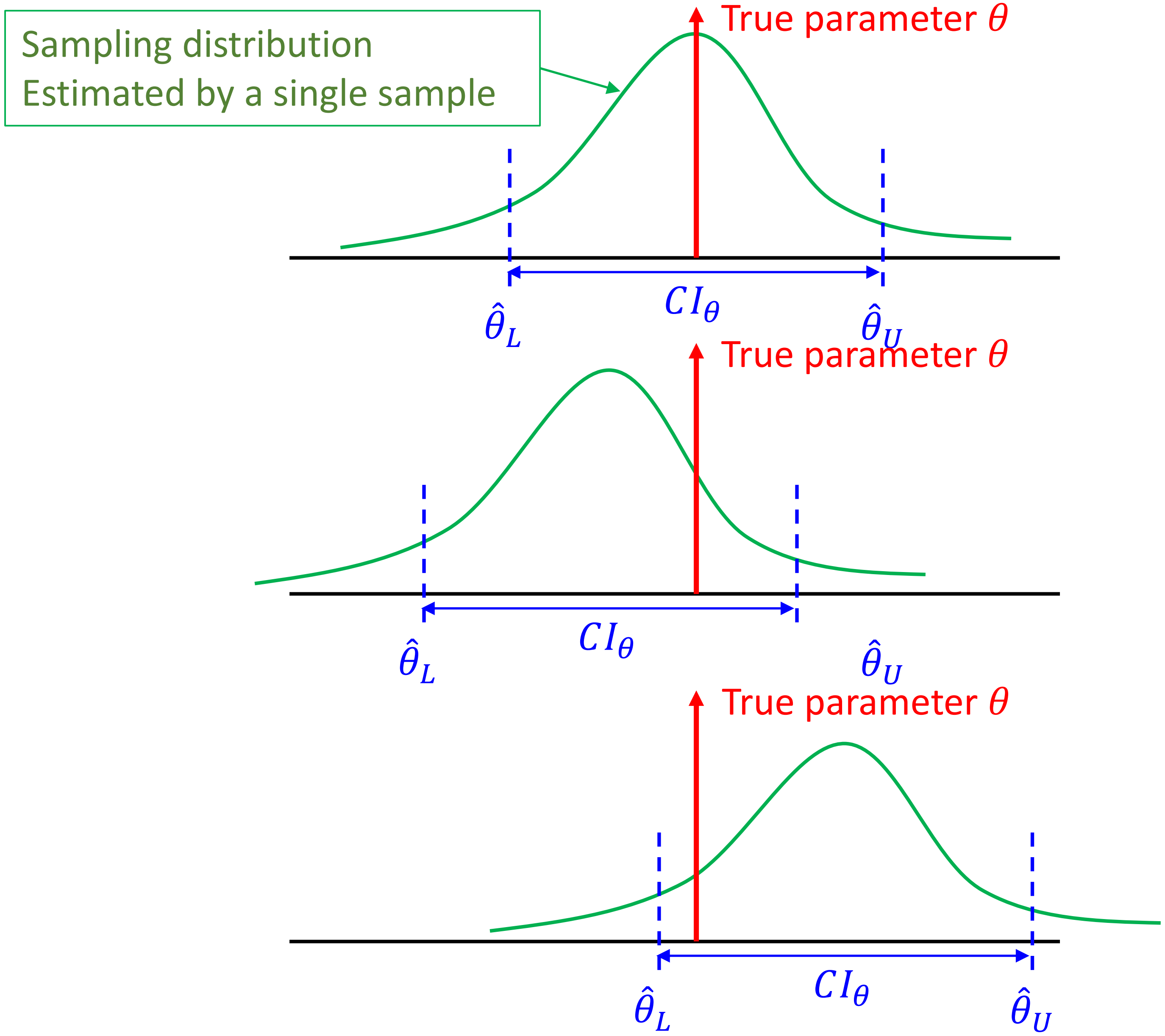
$$\sigma_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1{,}440{,}000}{100} + \frac{1{,}960{,}000}{100}} = 184.4\ miles.$$

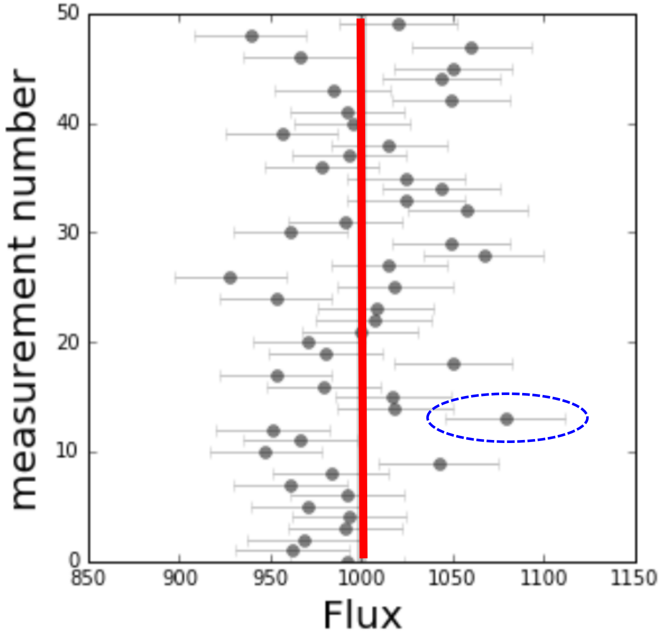
Consequently, we estimate the difference in mean wear to be 1300 miles, and we expect the error of estimation to be less than $2\sigma_{\bar{Y}_1-\bar{Y}_2}$, or 368.8 miles, with a probability of approximately 0.95.

Motivation

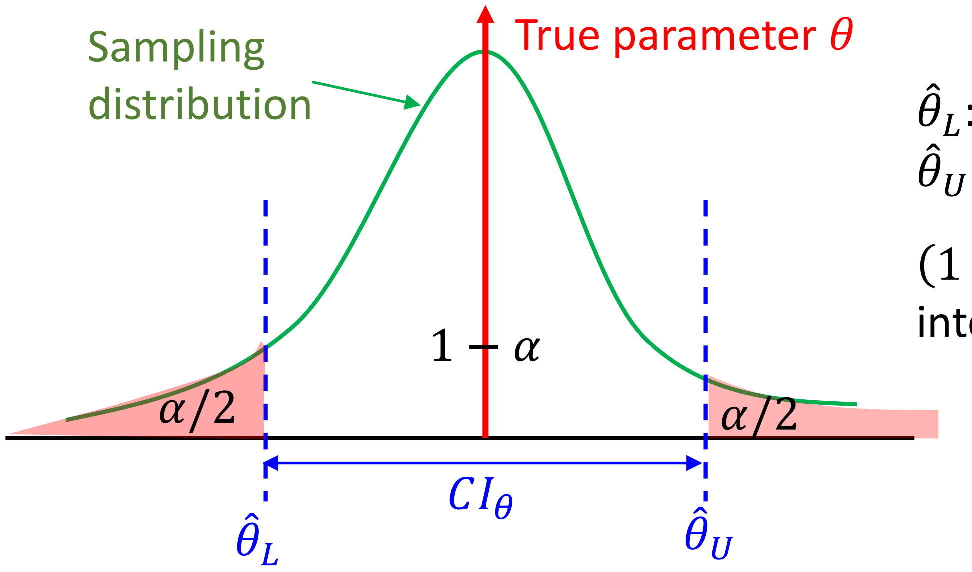
- An interval estimator is a rule specifying the method for using the sample measurements to calculate two numbers that form the endpoints of the interval.
- Ideally, the resulting interval will have two properties:
  - First, it will contain the target parameter θ;
  - Second, it will be relatively narrow.
- One or both of the end points of the interval, **being functions of the sample measurements (*Random variables*)**, will vary randomly from sample to sample.
  - Thus, the length and location of the interval are random quantities, and we cannot be certain that the (fixed) target parameter θ will fall between the endpoints of any single interval calculated from a single sample.
  - This being the case, our objective is to find an interval estimator capable of generating **narrow intervals that have a high probability of enclosing θ**.

Confidence Interval





Confidence Interval



$\hat{\theta}_L$: *The lower confidence li*mit, which is a random (function of a random samples)
$\hat{\theta}_U$: *The upper confidence limit*, which is a random (function of a random samples)

$(1-\alpha)$: confidence coefficient, the probability that a (random) confidence interval will enclose θ (a fixed quantity) is called the confidence coefficient

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_L\right) = 1-\alpha$$

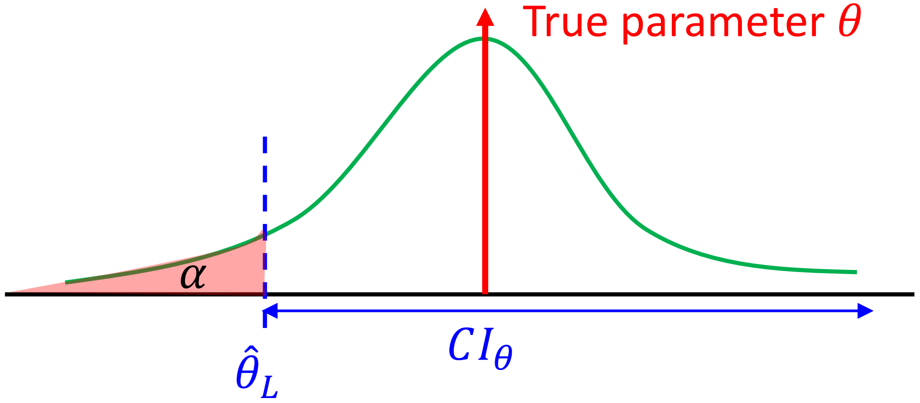
- *"There is a $(1-\alpha)$ % probability that when I compute the confidence interval (CI) from a current data sample, the computed CI contains θ*

  → From current data set, We can only say that $\theta \in \mathrm{CI}$ or $\theta \notin \mathrm{CI}$

- From a practical point of view, the confidence coefficient identifies the fraction of the time, **in repeated sampling**, that the intervals constructed will contain the target parameter θ.
  - If the confidence coefficient is high, we can be highly confident that any confidence interval, **constructed by using the results from a single sample**, will enclose θ.

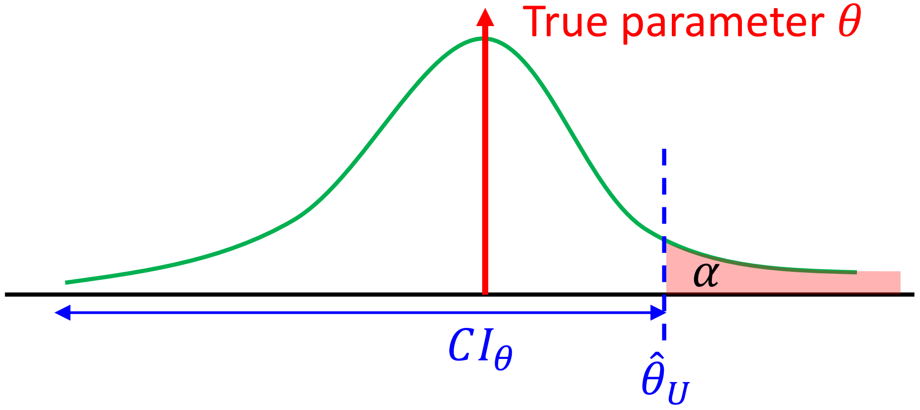
Confidence Interval

- An lower one-sided confidence interval



$$P(\hat{\theta}_L \le \theta) = 1 - \alpha$$

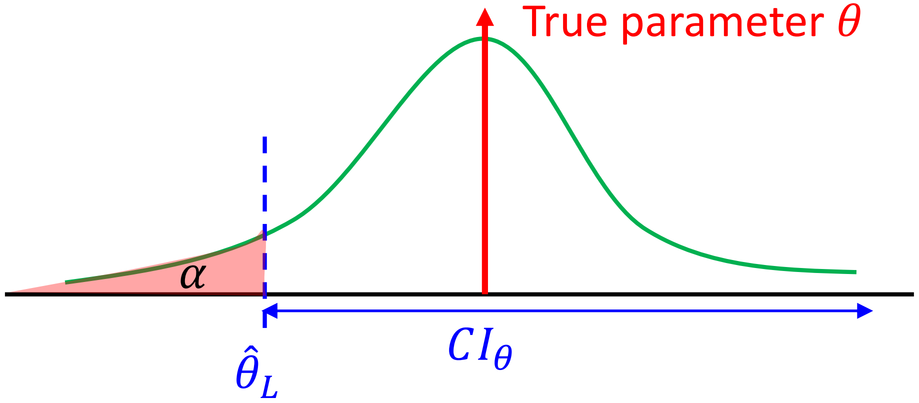
- An upper one-sided confidence interval



$$P(\theta \le \hat{\theta}_U) = 1 - \alpha$$

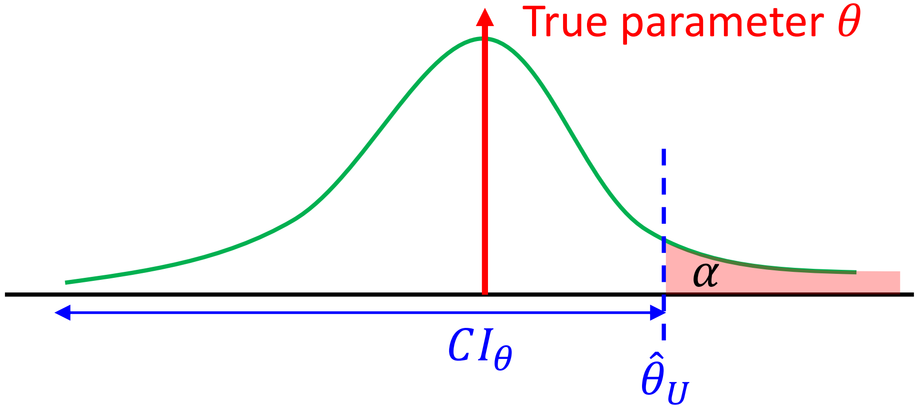
Confidence Interval

- An lower one-sided confidence interval



$$P(\hat{\theta}_L \le \theta) = 1 - \alpha$$

- An upper one-sided confidence interval



$$P(\theta \le \hat{\theta}_U) = 1 - \alpha$$

Pivotal Method for Estimating Confidence Interval

- This method depends on finding a pivotal quantity that possesses two characteristics
  - ✓ It is a function of the sample measurements and the unknown parameter θ, where θ is the only unknown quantity.
  - ✓ Its probability distribution does not depend on the parameter θ.

$$
\begin{aligned}
& P(a \leq Y \leq b) = .7 \\
\Rightarrow\ & P(ca \leq cY \leq cb) = .7 \\
\Rightarrow\ & P(a + d \leq Y + d \leq b + d) = .7
\end{aligned}
$$

- That is, the probability of the event $(a \leq Y \leq b)$ is unaffected by a change of scale or a translation of Y.
- Thus, if we know the probability distribution of a pivotal quantity, we may be able to use operations like these to form the desired interval estimator.

Example

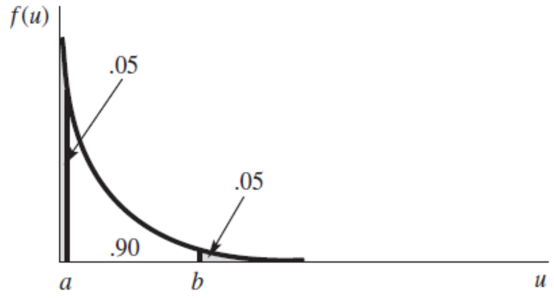
EXAMPLE 8.4

Suppose that we are to obtain a single observation Y from an exponential distribution with mean θ. Use Y to form a confidence interval for θ with confidence coefficient 0.90.

Example

SOLUTION 8.4

FIGURE 8.5
Density function for
U, Example 8.4



The probability density function for Y is given by

$$f(y) = \begin{cases} \left(\frac{1}{\theta}\right) e^{-y/\theta}, & y \geq 0 \\ 0, & elsewhere \end{cases}$$

By the transformation method we can see that $U = Y/\theta$ has the exponential density function given by

$$f_U(u) = \begin{cases} e^{-u}, & u > 0 \\ 0, & elsewhere \end{cases}$$

$U = Y/\theta$ is a function of Y (the sample measurement) and θ, and the distribution of U does not depend on θ. Thus, we can use $U = Y/\theta$ as a pivotal quantity. Because we want an interval estimator with confidence coefficient equal to 0.90, we find two numbers a and b such that

$$P(a \leq U \leq b) = 0.90.$$

Example

SOLUTION 8.4 (Cont.)

One way to do this is to choose a and b to satisfy

$$p(U < a) = \int_0^a e^{-u} du = 0.05, \qquad p(U > b) = \int_b^\infty e^{-u} du = 0.05$$

Thus, $a = 0.051,\ b = 2.996$.

$$0.90 = P(0.051 \leq U \leq 2.996) = p(0.051 \leq \frac{Y}{\theta} \leq 2.996)$$

Because we seek an interval estimator for θ, let us manipulate the inequalities describing the event to isolate θ in the middle. Y has an exponential distribution, so $p(Y > 0) = 1$, and we maintain the direction of the inequalities if we divide through by Y . Thus, we conclude that

$$0.90 = P(\frac{Y}{2.996} \leq \theta \leq \frac{Y}{0.051})$$

We know that limits of the form $(Y/2.996, Y/0.051)$ will include the true (unknown) values of θ for 90% of the values of Y we would obtain by repeatedly sampling from this exponential distribution.

Example

EXAMPLE 8.5

Suppose that we take a sample of size $n = 1$ from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ.

Example

SOLUTION 8.5

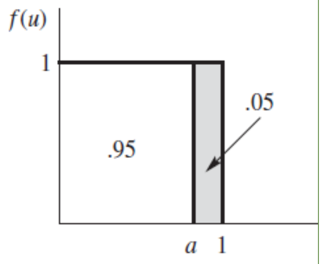
Because Y is uniform on $[0, \theta]$, the methods of Chapter 6 can be used to show that $U = Y/\theta$ is uniformly distributed over $[0, 1]$. That is,

$$f_U(u) = \begin{cases} 1, & 0 \le u \le 1 \\ 0, & elsewhere \end{cases}$$

Again, we see that U satisfies the requirements of a pivotal quantity. Because we seek a 95% lower confidence limit

for θ, let us determine the value for a so that $P(U \le a) = 0.95$. Thus, $a = 0.95$.

FIGURE 8.6 Density function for U, Example 8.5



$$p(U \le 0.95) = p\left(\frac{Y}{\theta} \le 0.95\right) = p\left(\frac{Y}{0.95} \le \theta\right) = 0.95.$$

We see that $Y/0.95$ is a lower confidence limit for θ, with confidence coefficient 0.95. Because any observed Y must be less than θ, it is intuitively reasonable to have the lower confidence limit for θ slightly larger than the observed value of Y .