ECE 340
Probabilistic Methods in Engineering
M/W 3-4:15

Lecture 21: Hypothesis Testing 1

Prof. Vince Calhoun

Homework

- Assignment 7: From the text: problems 7.2, 7.18, 7.21, 7.22, 7.25
- Due date: Mon, Apr 28 at the beginning of class

Parameter estimation

Point estimation

A parameter is a point estimation (numeric fact) of a population.

A statistic is a point estimation (numeric fact) of a sample.

$\Rightarrow$ A statistic is used to estimate a parameter.

Common point estimation of a parameter: μ is estimated by $\bar{X}$, σ^2 is estimated by S^2.

There may be different point estimations for a parameter.

How do we decide which point estimation is best for estimating a particular population parameter?

$\Rightarrow$Point estimation must be "close" to the true value of the unknown parameter being estimated.

[bias $\approx$ 0 (or = 0)]

if $E(\hat{\theta}) = \theta$, then $\hat{\theta}$ (statistic) is an unbiased estimation of θ (parameter).

$\Rightarrow$ $E(\hat{\theta}) = \theta$ says that the expected/ mean/ average value of the test statistic is the parameter.

Confidence Interval

Confidence Interval (CI) estimation for the mean
In many cases, a point estimate is not enough, a range of allowable values is better.
An interval of the form: Lower (l) $\leq \mu \leq$ Upper (u) may be useful.
To construct a CI for a parameter μ, we need to calculate 2 statistics l and u such that: $P(l \leq \mu \leq u) = 1 - \alpha$.

This interval is a 100(1- α)% CI for parameter.
l and u are lower and upper confidence limits.
1-α is called the confidence coefficient.
α = level of significance for Type I error (rejecting valid hypotheses).

Confidence Interval

- **Interpreting a confidence interval**
  μ is covered by interval with confidence 100(1-α)%.
- **If many samples are taken and a 100(1- α)% CI is calculated for each, then 100(1-α)% of them will contain/ cover the true value for μ.**
- **Note: the larger (wider) a CI, the more confident we are that the interval contains the true value of μ.**
- **But, the longer it is, the less we know about μ, due to variability or uncertainty $\rightarrow$ need to balance**

Confidence Interval

Confidence interval on mean, variance known

If: random sample of size n: $X_1, \ldots, X_n$
$X_i \sim N(\mu, \sigma^2)$ and $\bar{X} \sim N(\mu, \sigma^2/n)$

Then the test statistic: $Z = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2 / n}}$ **~ N(0, 1) by CLT**

With a CI, we want some range on μ, $P[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}] =$ α,

$$P[-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq Z_{\alpha/2}] = 1 - \alpha$$

$\rightarrow$ probability test statistic between 2 points is 1- α

Confidence Interval

P[-Z$_{\alpha/2}$ $\sigma/\sqrt{n} \leq \bar{X} - \mu \leq$ Z$_{\alpha/2}$ $\sigma/\sqrt{n}$] = 1- α $\rightarrow$ want a range on μ

P[- $\bar{X}$ + (-Z$_{\alpha/2}$ $\sigma/\sqrt{n}$)$\leq$ -μ $\leq$ - $\bar{X}$ + (Z$_{\alpha/2}$ $\sigma/\sqrt{n}$)] = 1- α

P[$\bar{X}$ + Z$_{\alpha/2}$ $\sigma/\sqrt{n}$ $\geq \mu \geq$ $\bar{X}$ - Z$_{\alpha/2}$ $\sigma/\sqrt{n}$] = 1- α

P[$\bar{X}$ - Z$_{\alpha/2}$ $\sigma/\sqrt{n}$ $\leq \mu \leq$ $\bar{X}$ + Z$_{\alpha/2}$ $\sigma/\sqrt{n}$] = 1- α

$\therefore$ a 100(1- α)% CI (2-sided) on μ is:

$\bar{X}$ - Z$_{\alpha/2}$ $\sigma/\sqrt{n}$ $\leq \mu \leq$ $\bar{X}$ + Z$_{\alpha/2}$ $\sigma/\sqrt{n}$

or $\bar{X}$ $\pm$ Z$_{\alpha/2}$ $\sigma/\sqrt{n}$

$\therefore$ A CI is a statistic $\pm$ (table value) x standard error.

Confidence Interval

CI on mean, variance unknown
Up to now, we have known σ. But typically we do not know, so what do we do?

1. **If $n \geq 30$, we can replace σ in the CI for the mean with the sample SD, S=** $\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(X_j - \mu\right)^2}$ **.**
2. **if $n < 30$, then if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$**

the the test statistic t = $\frac{\overline{X} - \mu}{S/\sqrt{n}}$ **~ t-distribution with (n-1) degrees of freedom.**

Confidence Interval

Get a CI for μ

P $[-t_{\alpha/2,\, n-1} \leq t \leq t_{\alpha/2,\, n-1}] = 1-\alpha$

P $\left[-t_{\alpha/2,\, n-1} \leq \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,\, n-1}\right] = 1-\alpha$

$\therefore$ **100(1-** α**)% CI on** μ **is:**

$\bar{X} - (t_{\alpha/2,\, n-1}\; S/\sqrt{n}) \leq \mu \leq \bar{X} + (t_{\alpha/2,\, n-1}\; S/\sqrt{n})$

or $\bar{X} \pm t_{\alpha/2,\, n-1}\; S/\sqrt{n}$

Normal Distribution

- **An example of converting to standard normal distribution is given by the data from a dry plasma etch study (Lynch and Markle, 1997). The data are in angstroms, from the “before” process improvement trial. The mean is 564.11, and the standard deviation is 10.747 angstroms.**

Run	1	2	3	4	5	6	7	8	9
Value	570.85	576.86	550.76	567.97	573.02	554.39	563.05	572.58	547.49
Z-value	0.6272	1.1864	-1.2422	0.3592	0.8291	-0.9044	-0.0986	0.7881	-1.5465

Confidence Interval

- **Continuing,**

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{(570.85 + 576.86 + \ldots + 547.49)}{9} = 564.11$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{9} (x_i - \bar{x})^2}{(9-1)}} = \sqrt{\frac{[(570.85 - 564.11)^2 + \ldots + (547.49 - 564.11)^2]}{8}} = 10.747$$

$$t_{8,0.025} = 2.306$$

$$95\% CI = \bar{x} \pm t_{\nu,\alpha/2} \frac{s}{\sqrt{n}} = 564.11 \pm (2.306)\left(\frac{10.747}{\sqrt{9}}\right) = \{555.85, 572.37\}$$

Hypothesis Tests

- Hypotheses defined
  - Used to infer population characteristics from observed data.
  - Hypothesis test: A series of procedures that allows us to make inferences about a population by analyzing samples
  - Key question: was the observed outcomes the result of chance variation, or was it an unusual event?
  - Hint: Frequency = Area = Probability

Hypothesis Tests

- Hypothesis: Definition of terms
  - Null hypothesis (H_0): Statement of no change or difference. This statement is tested directly, and we either reject H_0 or we do not reject H_0
  - Alternative hypothesis (H_1): The statement that must be true if H_0 is rejected.

Hypothesis Tests

- Definition of terms
  - Type I error: The mistake of rejecting H_0 when it is true.
  - Type II error: The mistake of failing to reject H_0 when it is false.
  - alpha risk (α):Probability of a type I error
  - beta risk (β): Probability of a type II error
  - Test statistic: sample value used in making decision about whether or not to reject H_0

Hypothesis Tests

- Definition of terms
  - Critical region: Area under the curve corresponding to test statistic that leads to rejection of H_0
  - Critical value: The value that separates the critical region from those values that do not lead to rejection of H_0
  - Significance level: The probability of rejecting H_0 when it is true
  - Degrees of freedom: Referred to as d.f. or ν, and = n - 1

Hypothesis Tests

- Definition of terms
  - Type I error: Producer's risk
  - Type II error: Consumer's risk
  - Set so type I is the more serious error type (taking action when none is required)
  - Levels for α and β must be established before the test is conducted

Hypothesis Tests

- Hypothesis: Definition of terms
  - Degree of freedom
    - Degree of freedom are a way of counting the information in an experiment. In other words, they relate to sample size. More specifically, d.f. = n – 1
    - A degree of freedom corresponds to the number of values that are free to vary in the sample. If you have a sample with 20 data points, each of the data points provides a distinct place of information. The data set is described completely by these 20 values. If you calculate the mean for this set of data, no new information is created because the mean was implied by all of the information in the 20 data points.

Hypothesis Tests

- **Hypothesis: Definition of terms**
  - **Degree of freedom**
    - **Once the mean is known, though, all of the information in the data set can be described with any 19 data points. The information in a 20^{th} data point is now redundant because the 20^{th} data points has lost the freedom to have any value besides the one imposed on it by the mean**
    - **We have one less than the total in our sample because a sample is at least one less than the total population.**

Hypothesis Tests

- If the population variance is unknown, use s of the sample to approximate population variance, since under central limit theorem, $s = \sigma$ when $n > 30$. Thus solve the problem as before, using s

- With smaller sample sizes, we have a different problem. But it is solved in the same manner. Instead of using the z distribution, we use the t distribution

Hypothesis Tests

- Using t distribution when:
  - Sample is small (<30)
  - Parent population is essentially normal
  - Population variance (σ) is unknown

  - As n decreases, variation within the sample increases, so distribution becomes flatter.

Methods to Test a Statistical Hypothesis

1) Calculate Test Statistics, check if it falls in expected value range, make conclusion based upon the result (hypothesis test).
2) Calculate confidence interval. If $H_0 : \mu = \mu_0$ falls in interval, fail to reject the null hypothesis H_0
3) P-value for an event
Reject H_0 if p-value $\leq \alpha$ = significant level.

If p-value < α, reject H_0
If p-value $\geq \alpha$, fail to reject

Methods to Test a Statistical Hypothesis

Calculate Test Statistics, Test if it falls in Critical Region (CR), make conclusion (hypothesis test).
If Test Statistic > CR, reject H_0

Calculate confidence interval. If μ_0 falls in interval, fail to reject the null hypothesis H_0

Calculate Test Statistic, Calculate p-value.
If p-value < α, reject H_0

Relationship Between Hypothesis Tests and Confidence Intervals

For a two-sided hypothesis test:

$H_0 : \mu = \mu_0$ $H_a : \mu \neq \mu_0$

Equivalent confidence interval is: (lower-limit, upper-limit)

If μ_0 is contained within the two-sided interval you will fail to reject H_0

If μ_0 is not contained within the two-sided interval you will reject H_0

Relationship Between Hypothesis Tests and Confidence Intervals

i. For an upper-tail test:
$H_0 : \mu = \mu_0$ $H_a : \mu > \mu_0$
Equivalent confidence interval is: (lower-limit, ∞)

Use the **lower bound** interval for comparison.

ii. For an lower-tail test:
$H_0 : \mu = \mu_0$ $H_a : \mu < \mu_0$
Equivalent confidence interval is: (∞, upper-limit)

Use the **upper bound** interval for comparison

Relationship between Hypothesis Tests and Confidence Intervals

- Using the data from the plasma etch study, can a true process mean of 530 angstroms be expected at a 95% confidence level?
- The 95% confidence interval (developed earlier in detail) runs from 555.85 to 572.37. Since 530 is not included in this interval, the null hypothesis of $\mu = 530$ is rejected.

Confidence Interval

An example: A dry plasma etching process has had a series of process variability improvements made. As part of a study by Lynch and Markle (1997), is it possible that the new process has significantly shifted from the process before improvements? If the 95% confidence interval does not contain the previous mean before changes, the process is considered to have significantly shifted. The old process mean is 564.108 angstroms per minute etch rate.

$$\bar{x}_{new} = 552.342 \ {}^{angstroms}\!/_{\min}$$

$$s_{new} = 2.3280 \ {}^{angstroms}\!/_{\min}$$

$$n_{new} = 9$$

$$t_{0.025, 8df} = 2.306$$

Do we expect that the process has shifted to the 95% confidence level?

95% CI = $\bar{X} \pm t_{\alpha/2,\, n-1}\, S/\sqrt{n}$ **= 552.342 ± 2.306*(2.3280/3.00)**

95% CI = {550.553, 554.131} $\therefore$ **Yes, the process mean has shifted.**

P - Values

- **The P – value is the smallest level of significance that leads to rejection of the null hypothesis with the given data. It is the probability attached to the value of the Z statistic developed in experimental conditions. It is dependent upon the type of test (two-sided, upper, or lower tail tests) selected to analyze data significance.**

Confidence Interval

Hypothesis test

Reject H_0 if p-value $\leq \alpha$ = significant level.

If p-value < α, reject H_0
If p-value $\geq \alpha$, fail to reject

Note: the larger (wider) a CI, the more confident we are that the interval contains then true value of μ.
But, the longer it is, the less we know about μ,
$\rightarrow$ need to balance

Groundwork for Inferential Statistics

- Recall that, our primary concern is to make inference about the population under study.
- Since we cannot study the entire population we rely on a subset of the population, called sample, to make inference.
- We saw how to take samples.
- Having taken the sample, how do we make inference on the population?

Basic Concepts

Definition

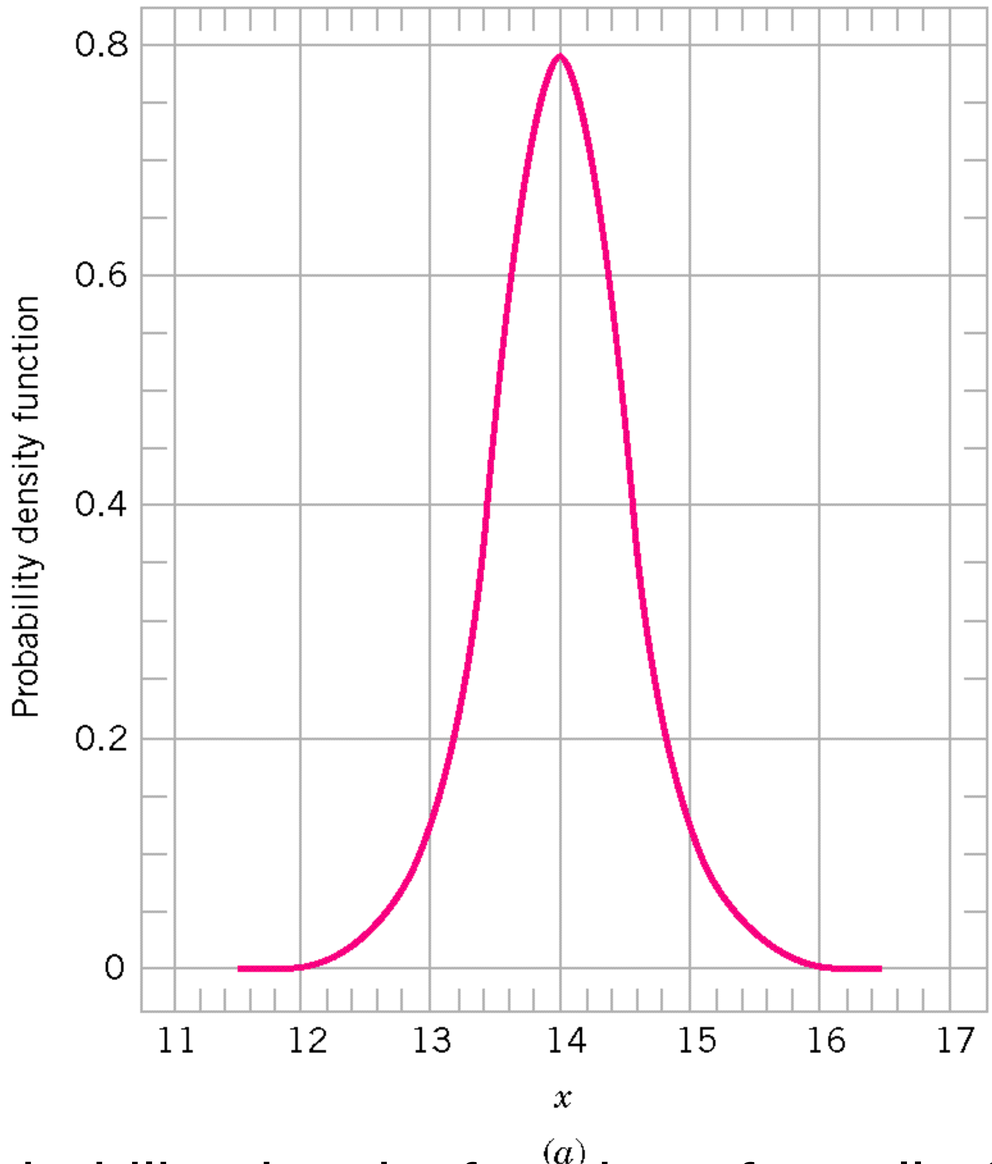
Independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution are called a **random sample.**

Definition

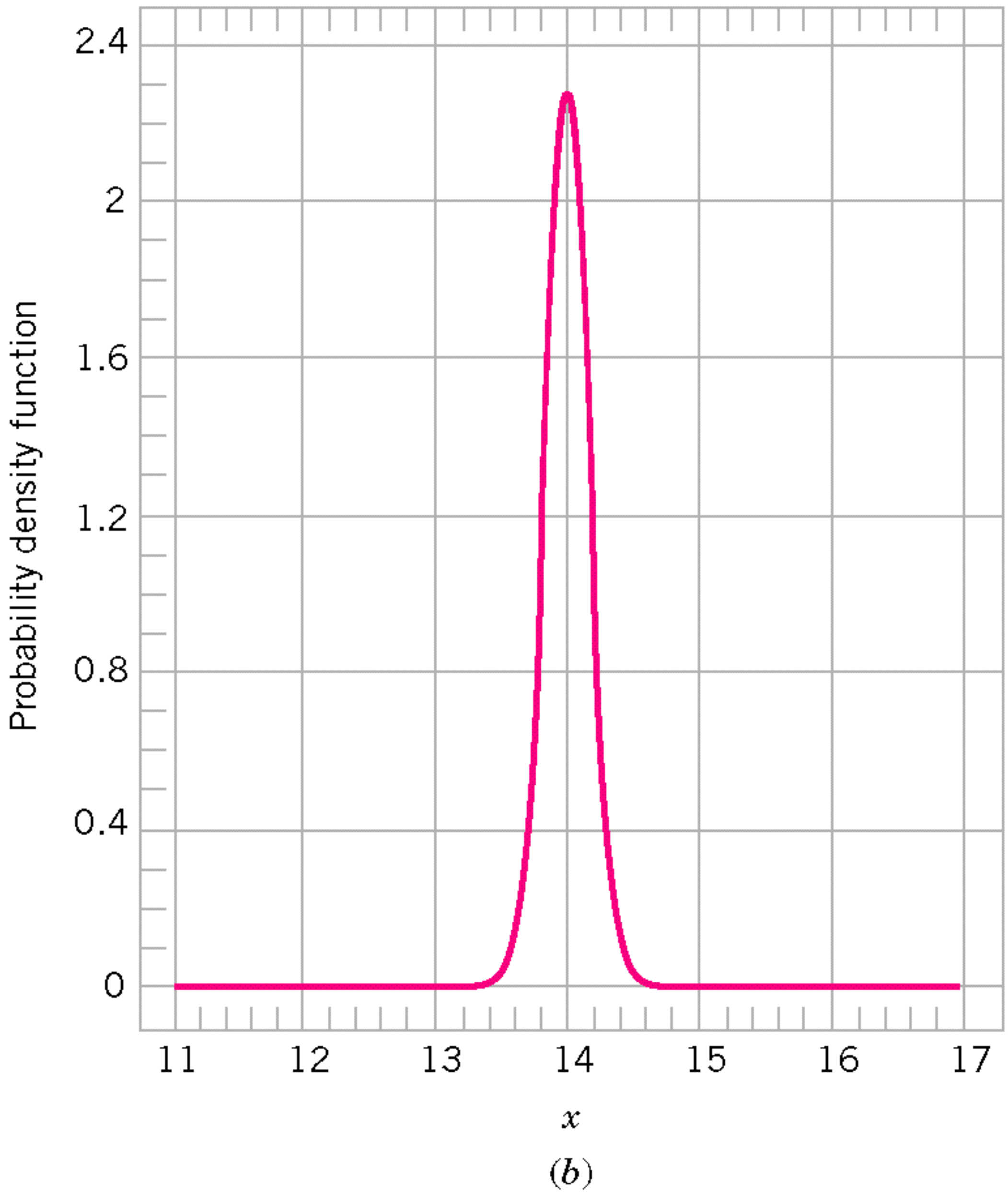
A **statistic** is a function of a random sample.

Definition

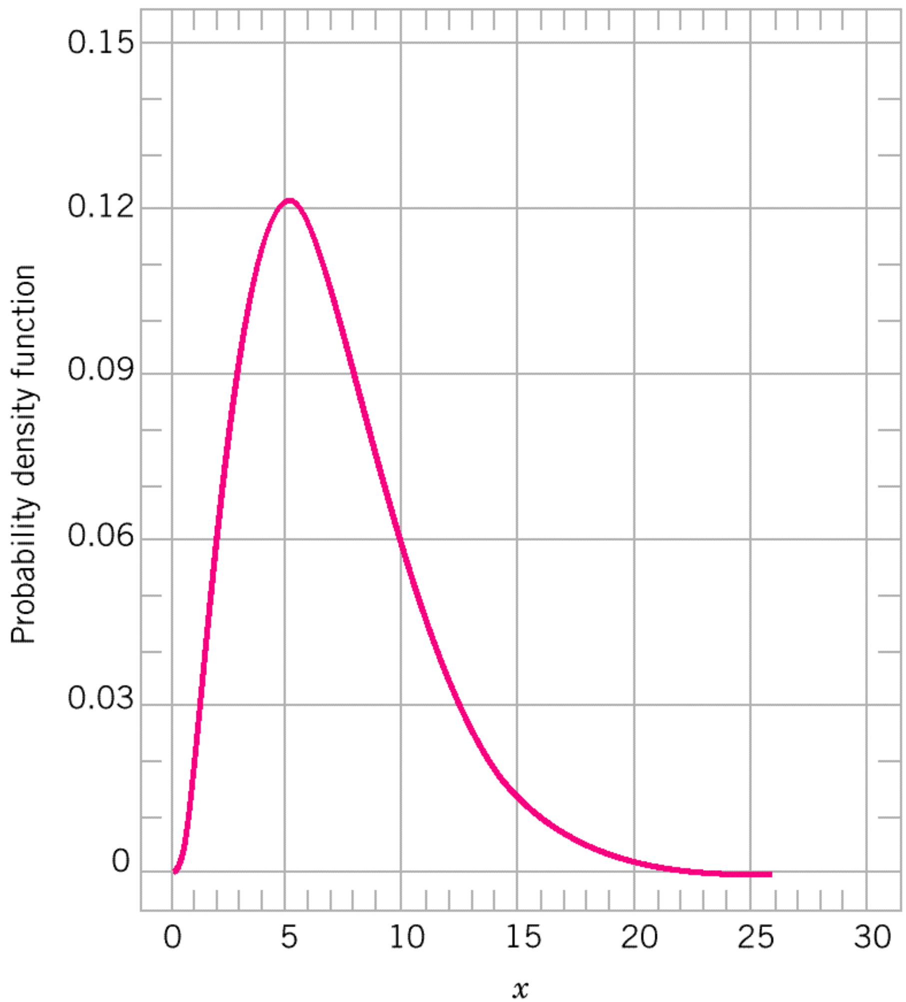
The probability distribution of a statistic is called its **sampling distribution.**



Probability density function of a pull-off force measurement



Probability density function of the average of 8 pull-off force measurements.



(c)

Probability density Probability density function
function of the sample variance of 8 pull-off force
measurements.

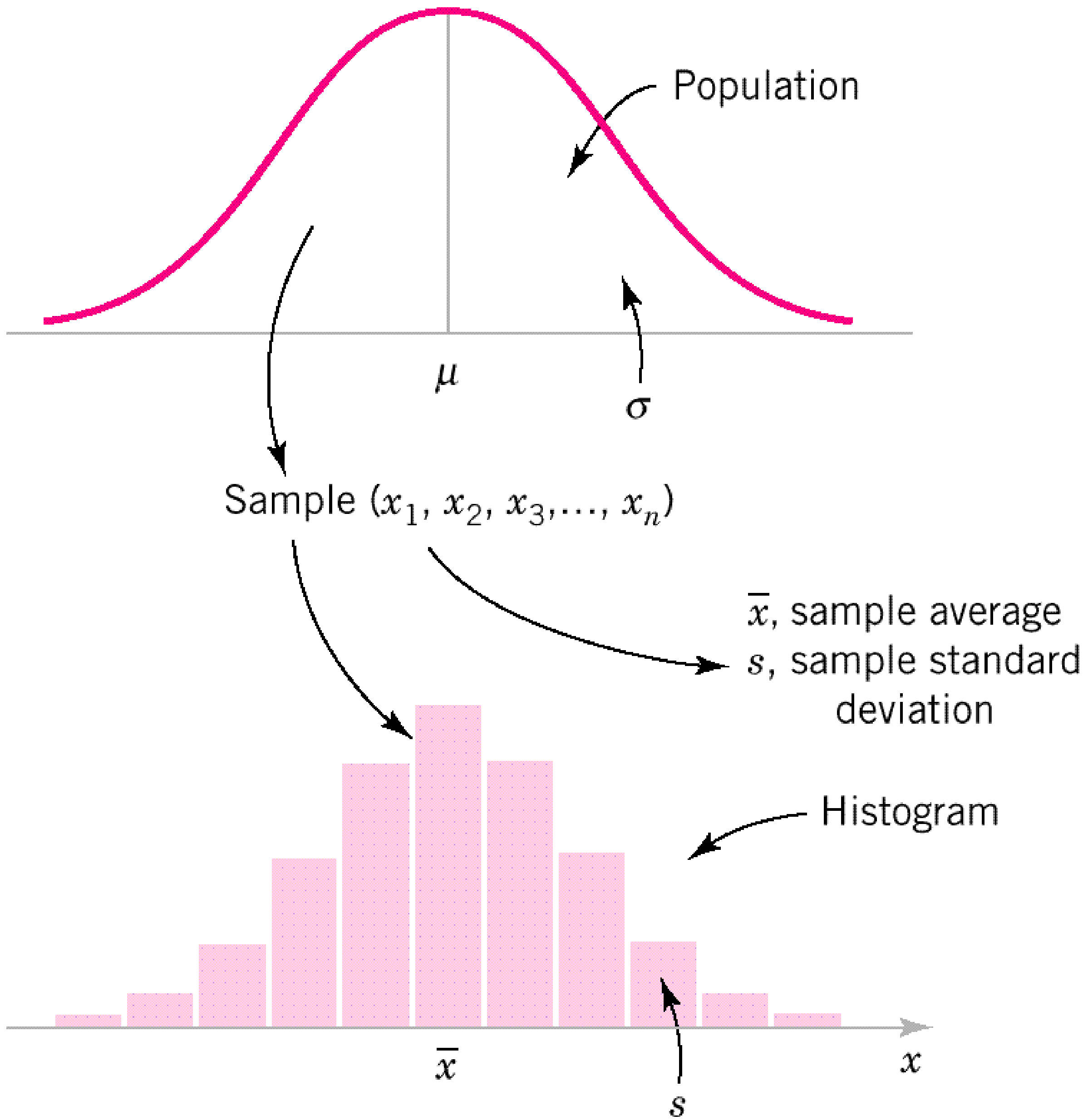


Figure 4-1 Relationship between a population and a sample.

Tests of Hypotheses

Definition

A **statistical hypothesis** is a statement about the parameters of one or more populations.

- **Two types of hypotheses: Null (H_0)and alternative (H_1)**

Table 4-1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	no error	type II error
Reject H_0	type I error	no error

Definition

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error.**

Definition

Failing to reject the null hypothesis when it is false is defined as a **type II error.**

Basic Ideas in Tests of Hypotheses

- Set up H_0 and H_1. For a one-sided case, make sure these are set correctly. Usually these are done such that type 1 error becomes "costly" error.
- Choose appropriate test statistic. This is usually based on the UMV estimator of the parameter under study.
- Set up the decision rule.
- Choose a random sample and make the decision.

Setting up H_o and H_1

- Suppose that the manufacturer of airbags for automobiles claims that the mean time to inflate airbag is no more than 0.1 second.
- Suppose that the “costly error” is to conclude erroneously that the mean time is < 0.1.
- How do we set up the hypotheses?

Definition

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

Definition

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0.

$$P = \begin{cases} 2[1 - \Phi(|z_0|)] & \text{for a two-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu \neq \mu_0 \\ 1 - \Phi(z_0) & \text{for an upper-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu > \mu_0 \\ \Phi(z_0) & \text{for a lower-tailed test: } H_0\text{: } \mu = \mu_0 \quad H_1\text{: } \mu < \mu_0 \end{cases} \qquad (4\text{-}22)$$

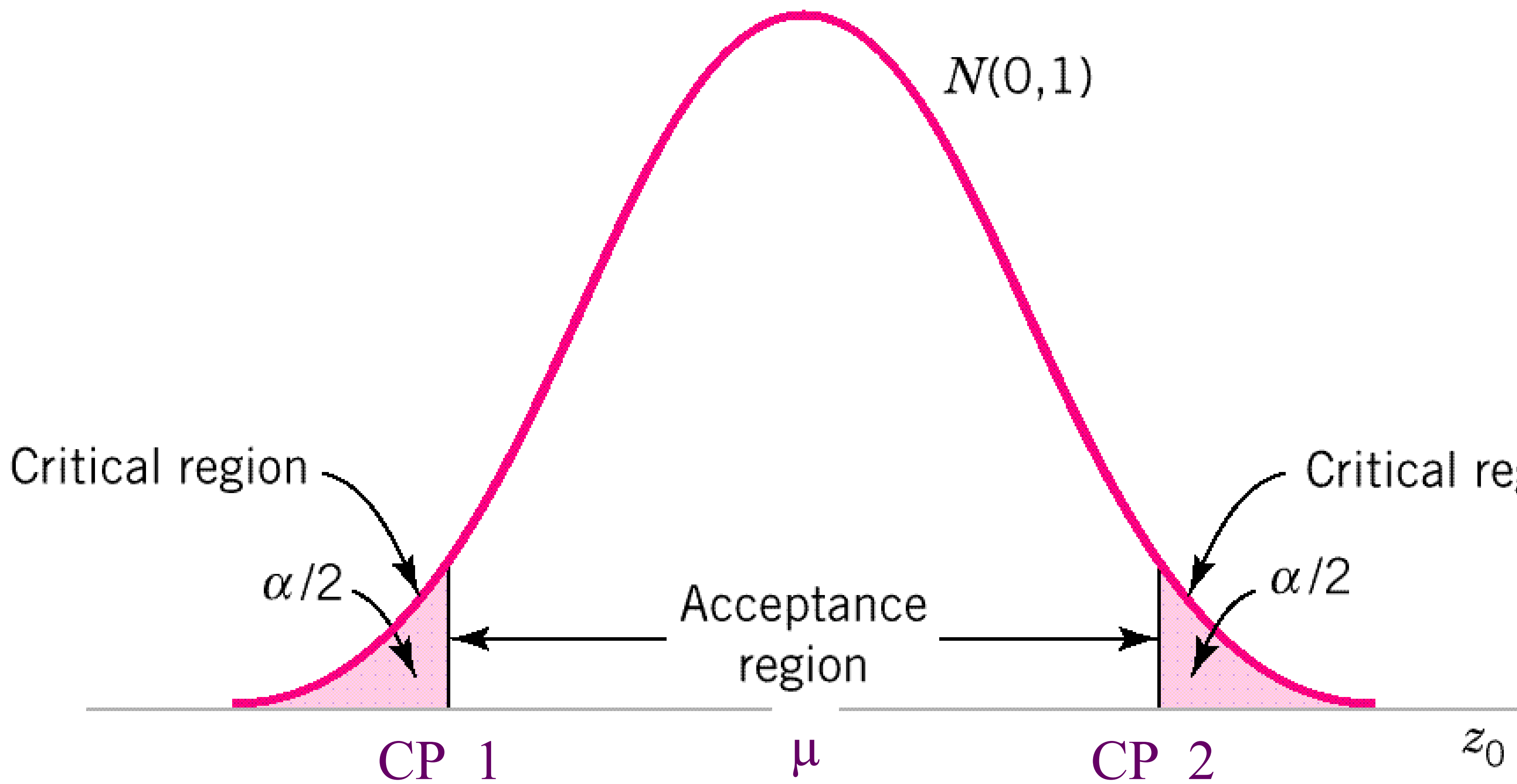


Figure 4-8 The distribution of Z_0 when H_0: $\mu = \mu_0$ is true, with critical region for H_1: $\mu \neq \mu_0$.

Example

- μ = Mean propellant burning rate (in cm/s).
- H_0:μ = 50 vs H_1:μ = 52.
- Two-sided hypotheses.
- A sample of n=10 observations is used to test the hypotheses.
- Suppose that we are given the decision rule.
- Question 1: Compute P(type 1 error)
- Question 2: Compute P(type 2 error when μ =52.

DECISION RULE

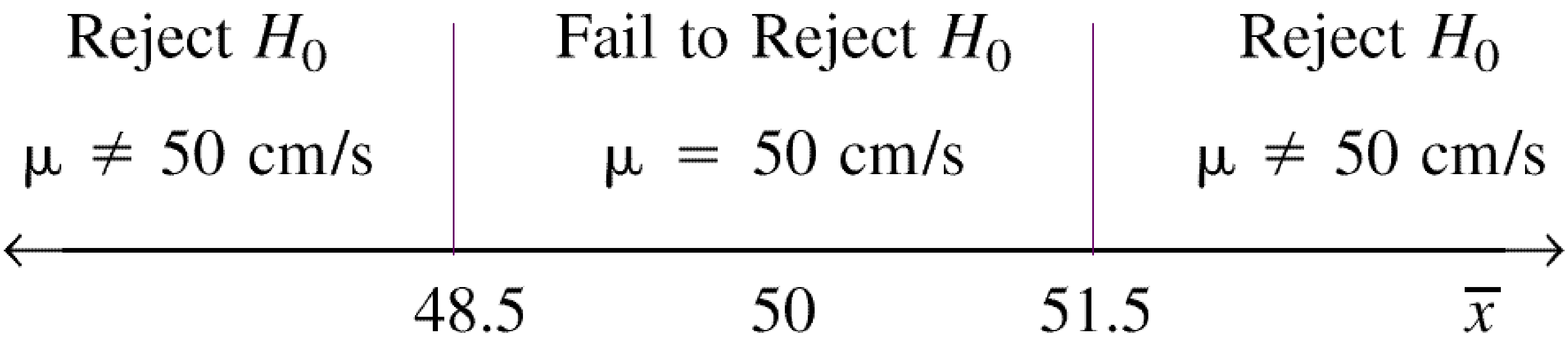


Figure 4-3 Decision criteria for testing H_0: $\mu = 50$ cm/s versus H_1: $\mu \neq 50$ cm/s.

Calculation of
P(type 1 error)

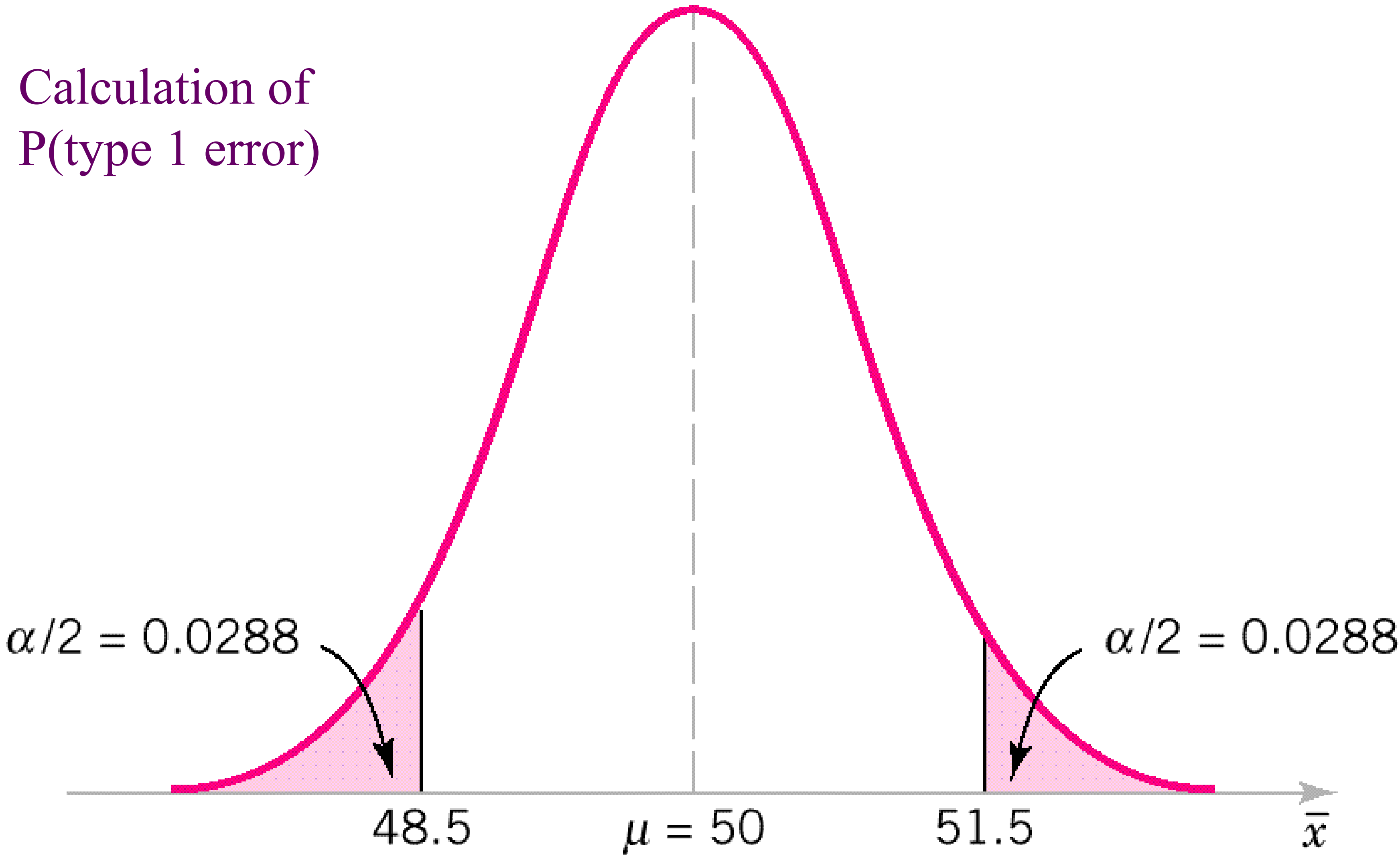


Figure 4-4 The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

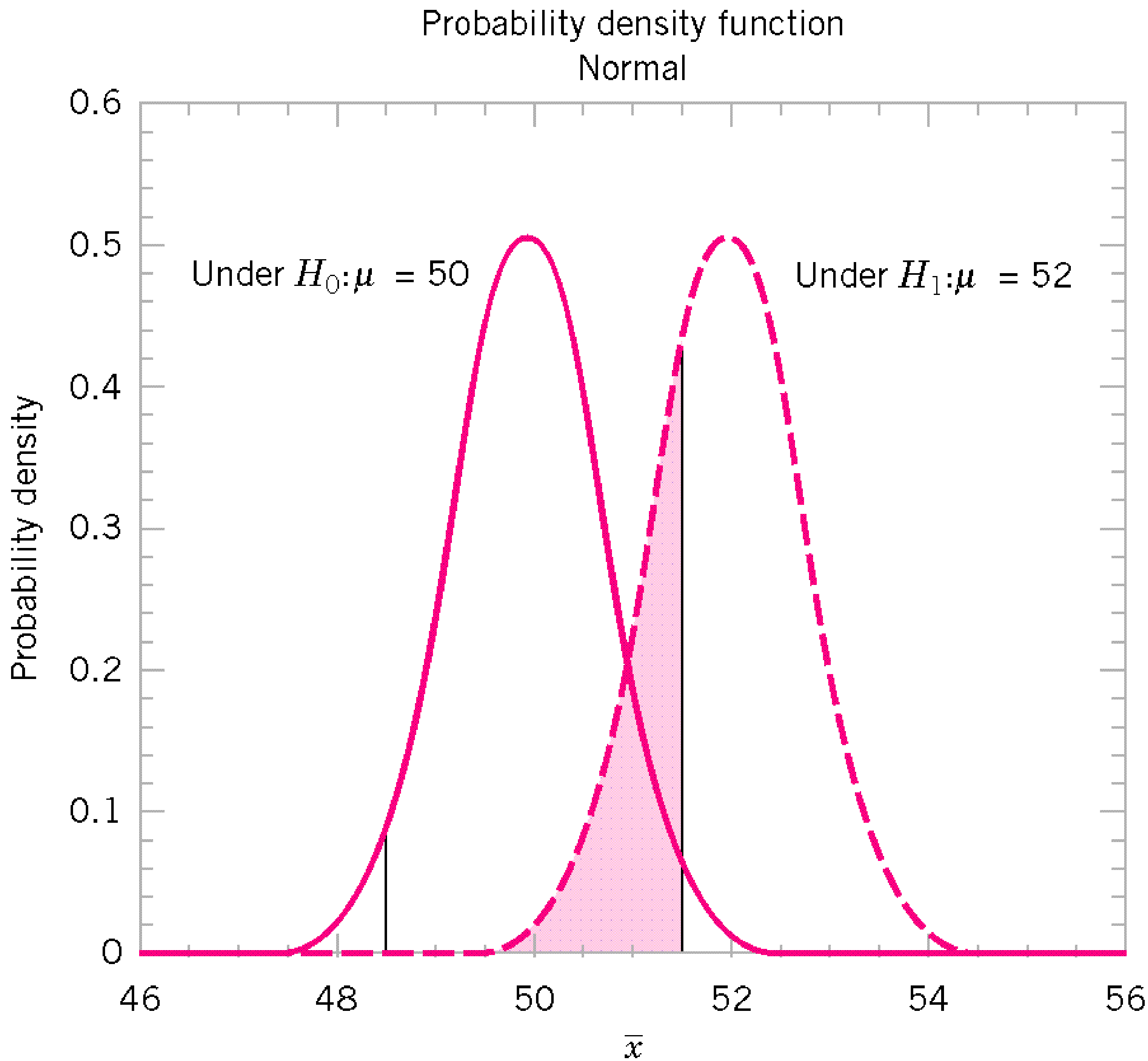
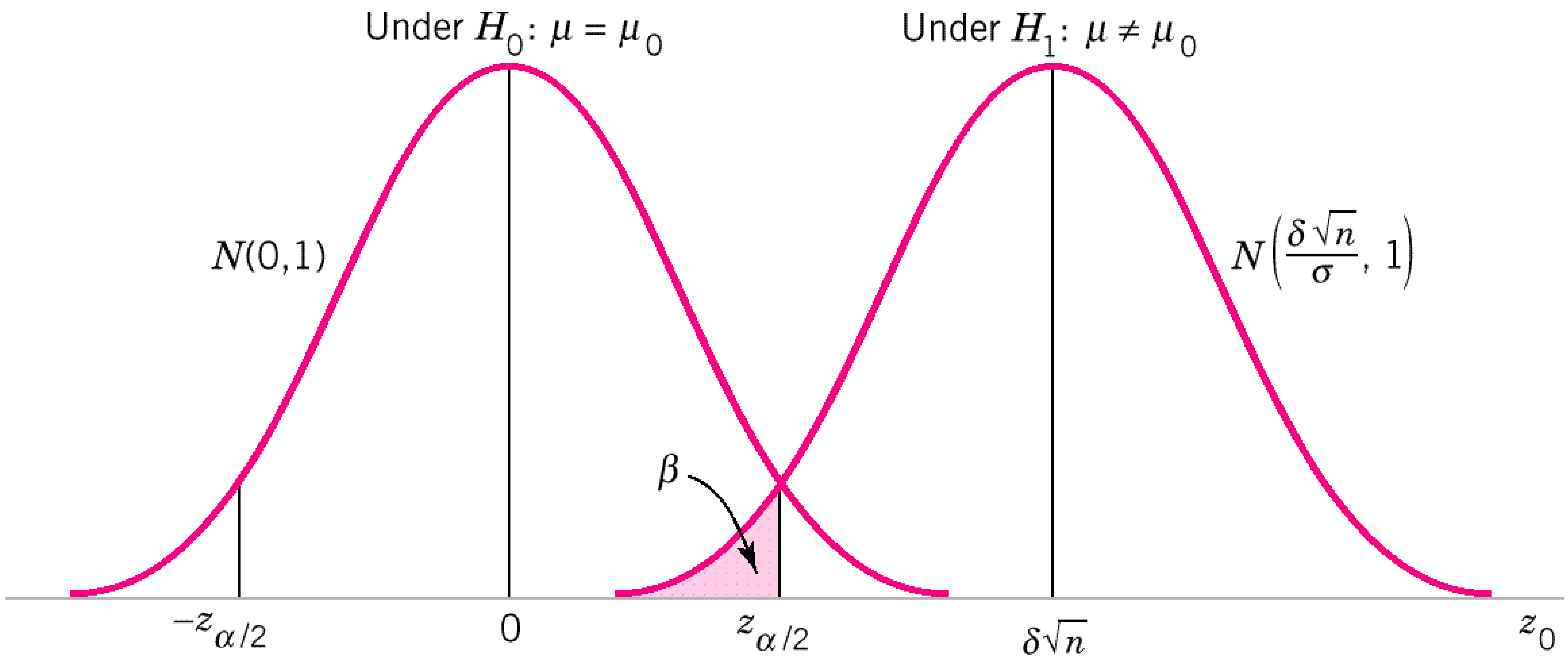


Figure 4-5 The probability of type II error when μ = 52 and n = 10.

Confidence Interval

- Recall point estimate for the parameter under study.
- For example, suppose that μ= mean tensile strength of a piece of wire.
- If a random sample of size 36 yielded a mean of 242.4psi.
  - Can we attach any confidence to this value?
- Answer: No! What do we do?

Sample Size Determination for a given δ



$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \quad (4\text{-}24)$$

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2} \quad (4\text{-}26)$$

where

$$\delta = \mu - \mu_0$$

Student's t-distribution

- Referring to HP example, we assumed that the population standard deviation was known (to be 10).
- However, in practice, it is usually unknown. Hence, we need to estimate it first. If the sample size is reasonably large ($n \geq 30$), we can still use the normal distribution for inferential part (as justified by the CLT).

Student’s t-distribution

- What happens if the sample is small (n < 30)?
- In this case we cannot use normal since the sample size is small and by using the sample standard deviation to estimate s, we bring in more variability into the picture and the appropriate distribution to use is the student's t-distribution.
- In 1908, William S.Gosset, a chemist working for a brewery company, under the pseudonym Student, first deduced this distribution

Student's t-distribution

- Suppose that X_1, X_2, …, X_n are n random samples from a normal distribution with mean μ and standard deviation s. Then the PDF of

$$T = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

- is given by

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \frac{1}{[(t^2/k)+1]^{(k+1)/2}}, -\infty < t < \infty,$$

$$\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx, \ \textit{for any positive number } k.$$

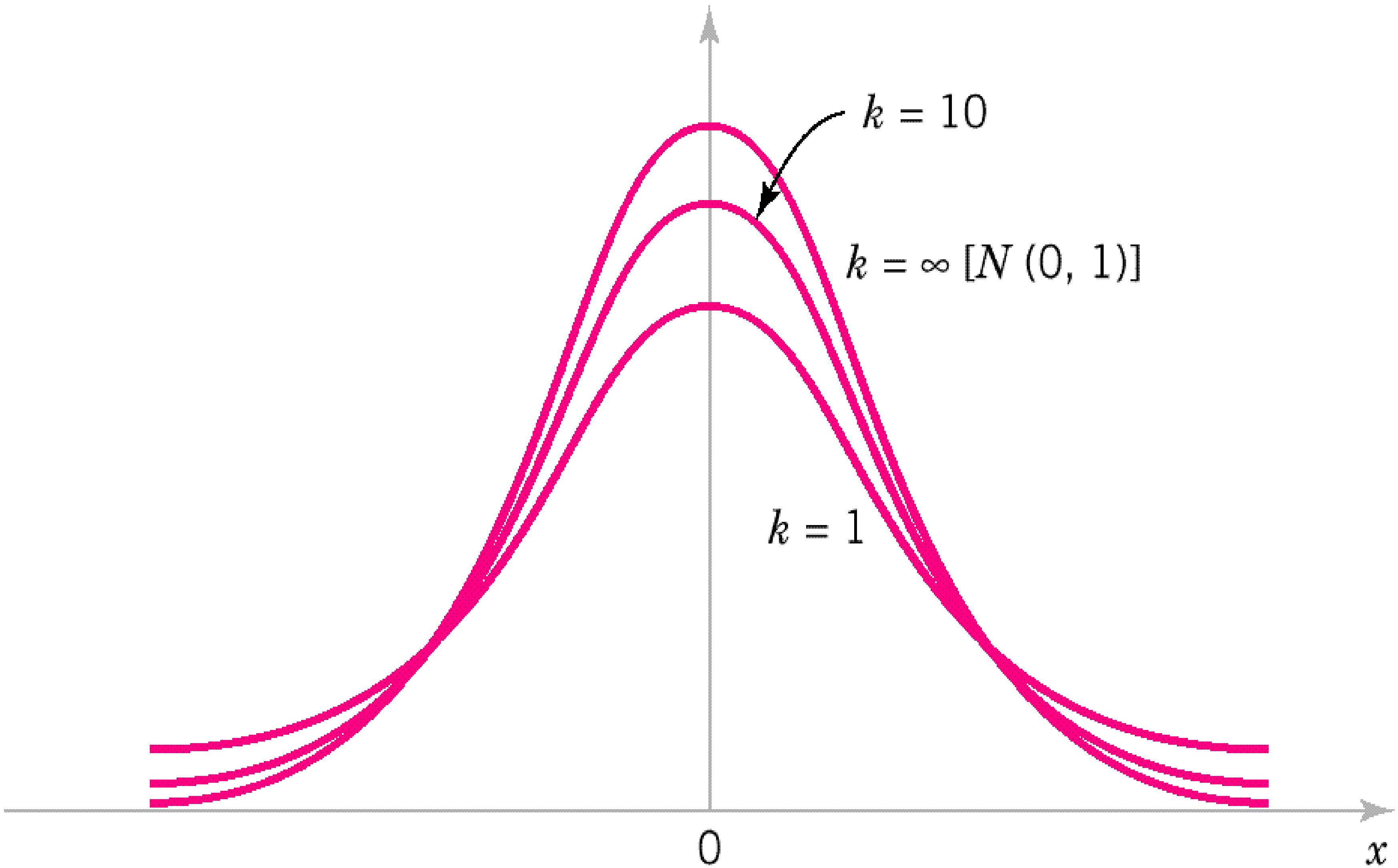


Figure 4-13 Probability density functions of several t distributions.

Student’s t-distribution (cont’d)

- Student’s t-distribution, like normal,
  - is bell-shaped. It depends on the sample size.
  - It is more spread than normal and approaches normal as n approaches infinity.