Unit #0 Basic Statistical Inference (Review?)

See notes linked in the course schedule

The Big Picture: Probability Vs. Statistics

- Let the probability of heads for a given coin be $p = 0.3$. What is the probability that, out of 10 flips, fewer than 3 land on heads?

- Suppose that we observe 10 flips of a coin with unknown probability p, and three of those flips land on heads. What is p?

The Big Picture: Probability Vs. Statistics

Probability is about the future.

Suppose you have an unfairly weighted coin that, if flipped, will result in "heads" 57% of the time. Suppose that **you are going** to flip this coin 10 times. What is the probability that you **will** see at least 7 heads? How many heads **do you expect** to see?

Did you catch all those future type words and phrases there?

Statistics, on the other hand, **is about the past**.

That is, suppose that you know next to nothing about the coin but you flipped it 10 times and observed the heads/tails data

$$H, H, T, T, T, H, H, T, H, H.$$

You can use statistics to attempt figure out (albeit with some uncertainty) whether or not the coin **was** fair in the first place and what that single flip heads probability **might have been**.

Do you see how statistics is about looking **back** to figure out what was going on with the coin?

Statistical Inference: Estimation (Maximum Likelihood Estimation)

The Big Picture

Let $X_1, X_2, \ldots, X_n$ be a random sample from the N(μ, 1) distribution. How would you estimate μ? Intuition says:

(a) How do we justify this intuition?

(b) How "good" is this estimator? What does "good" even mean?

(c) Can we do better?

(d) How would you go about estimating parameters with a less obvious interpretation, such as the α from the $\Gamma(\alpha,\beta)$ distribution?

Estimator, Estimate, and Statistic

Definition: Suppose that θ is a fixed parameter to be estimated. Then an *estimator* of θ is a rule used for calculating an estimate of θ from a sample $X_1, \ldots, X_n$. We usually denote an estimator as $\widehat{\theta}$, or more precisely, as $\widehat{\theta}(X_1, ..., X_n)$, to emphasize that our estimator is a function of a pre-observed sample.

Definition: An *estimate* is used to refer to a particular observation of $\widehat{\theta}$; that is, $\widehat{\theta}(x_1, ..., x_n)$.

Examples:

Estimator, Estimate, and Statistic

Definition: A *statistic* is a function of the data:

$$T = t(X_1, ..., X_n) = t(\mathbf{X})$$

Note: estimators are statistics.

How do we come up with an estimator for a given parameter?

Maximum Likelihood Estimation

Motivating Example: Suppose you have an unfair coin where the parameter of interest, p = P (“Heads”), is known to be one of 0.2, 0.3, or 0.8.

Let’s suppose that we toss the coin twice and use the results to try to estimate p.

What is the distribution modeling the data?

Maximum Likelihood Estimation

Motivating Example: Suppose you have an unfair coin where the parameter of interest, *p* = P ("Heads"), is known to be one of 0.2, 0.3, or 0.8. Let's suppose that we toss the coin twice and use the results to try to estimate *p*.

		(x_1, x_2)			
		(0,0)	(0,1)	(1,0)	(1,1)
	0.2	0.64	0.16	0.16	0.04
p	0.3	0.49	0.21	0.21	0.09
	0.8	0.04	0.16	0.16	0.64

Maximum Likelihood Estimation

More generally, the maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the likelihood function.

Definition: A *likelihood function,* denoted by *L(θ),* is defined as any function proportional to the joint pdf ________, but thought of as a function of *θ*.

Definition: A *log likelihood function,* denoted by $\ell(\theta)$, is the log of a likelihood function.

Often, it is more convenient to maximize the log likelihood (and the maximizer will always be the same!).

Maximum Likelihood Estimation

Example: Suppose we have a coin with unknown probability of heads, p, where $\{p : 0 \leq p \leq 1\}$ (this is called the parameter space of p). We flip the coin n times and record the number of heads. Find the MLE for p.

Maximum Likelihood Estimation

Example: Let X_1, X_2,…,X_n be a random sample with pdf ___________________________. Find the MLE of _______.

Maximum Likelihood Estimation

Example: (Two parameter normal case)

Maximum Likelihood Estimation

Example:

Estimators and Their Distributions

We use *estimators* to summarize our i.i.d. sample. Any estimator, including the *sample mean* ________ is a random variable (since it is based on a random sample).

This means that ______ has a distribution of it's own, which is referred to as **sampling distribution of the sample mean.** This sampling distribution depends on:

The standard deviation of this distribution is called **the standard error of the estimator.**

Distribution of the Sample Mean

Let X_1, X_2, . . . , X_n be a random sample from a distribution with mean value and standard deviation . Then:

The standard deviation of the sample mean is:

This is also called the standard error of the mean.

Distribution of the Sample Mean

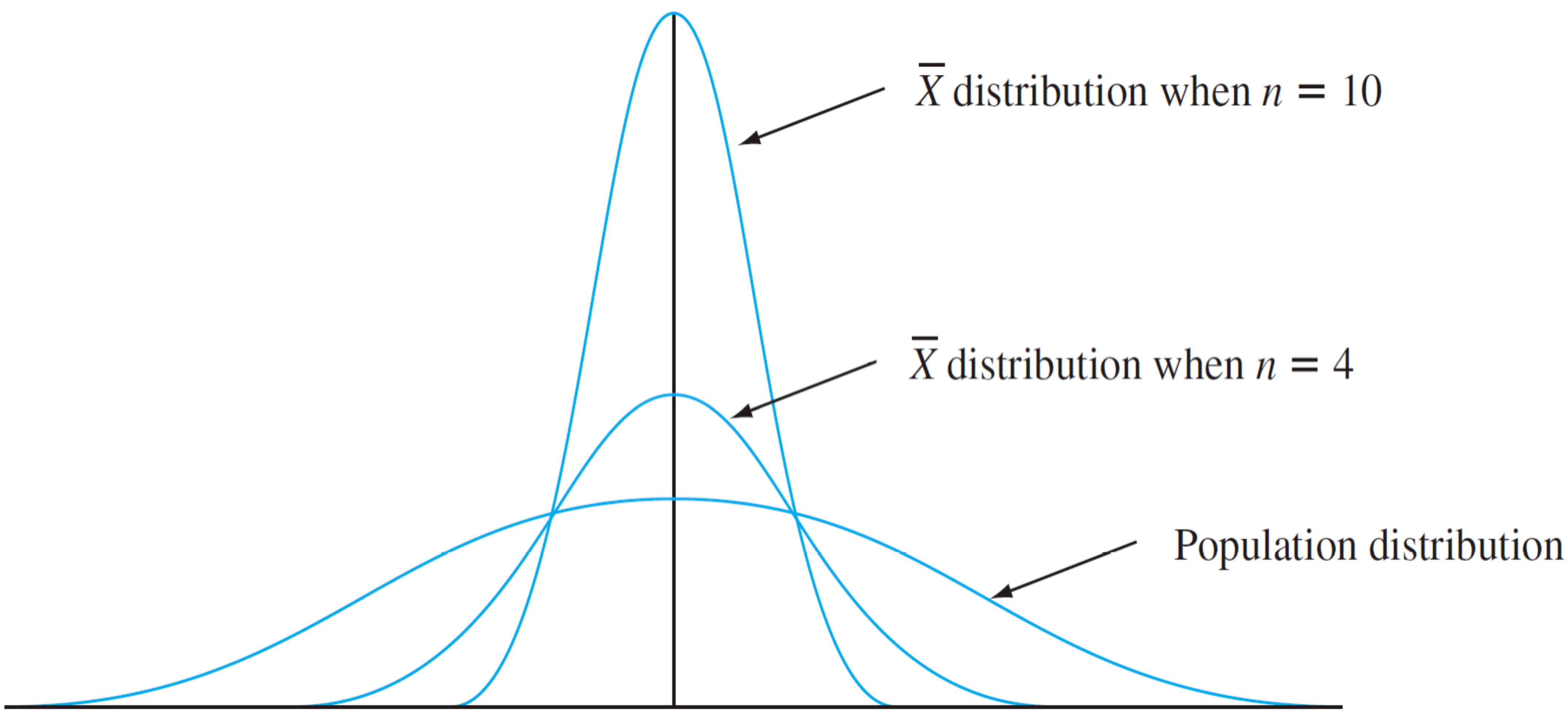
Great, but what is the *distribution* of the sample mean?

Distribution of the Sample Mean (Normal Population)

Proposition:

We know everything there is to know about the distribution of the sample mean when the population distribution is normal.

Distribution of the Sample Mean (Normal Population)



The Central Limit Theorem

But what if the underlying distribution of the X_i's is not normal?

The Central Limit Theorem

Important: When the population distribution is non-normal, averaging produces a distribution more bell-shaped than the one being sampled.

A reasonable conjecture is that if n is large, a suitable normal curve will approximate the actual distribution of the sample mean.

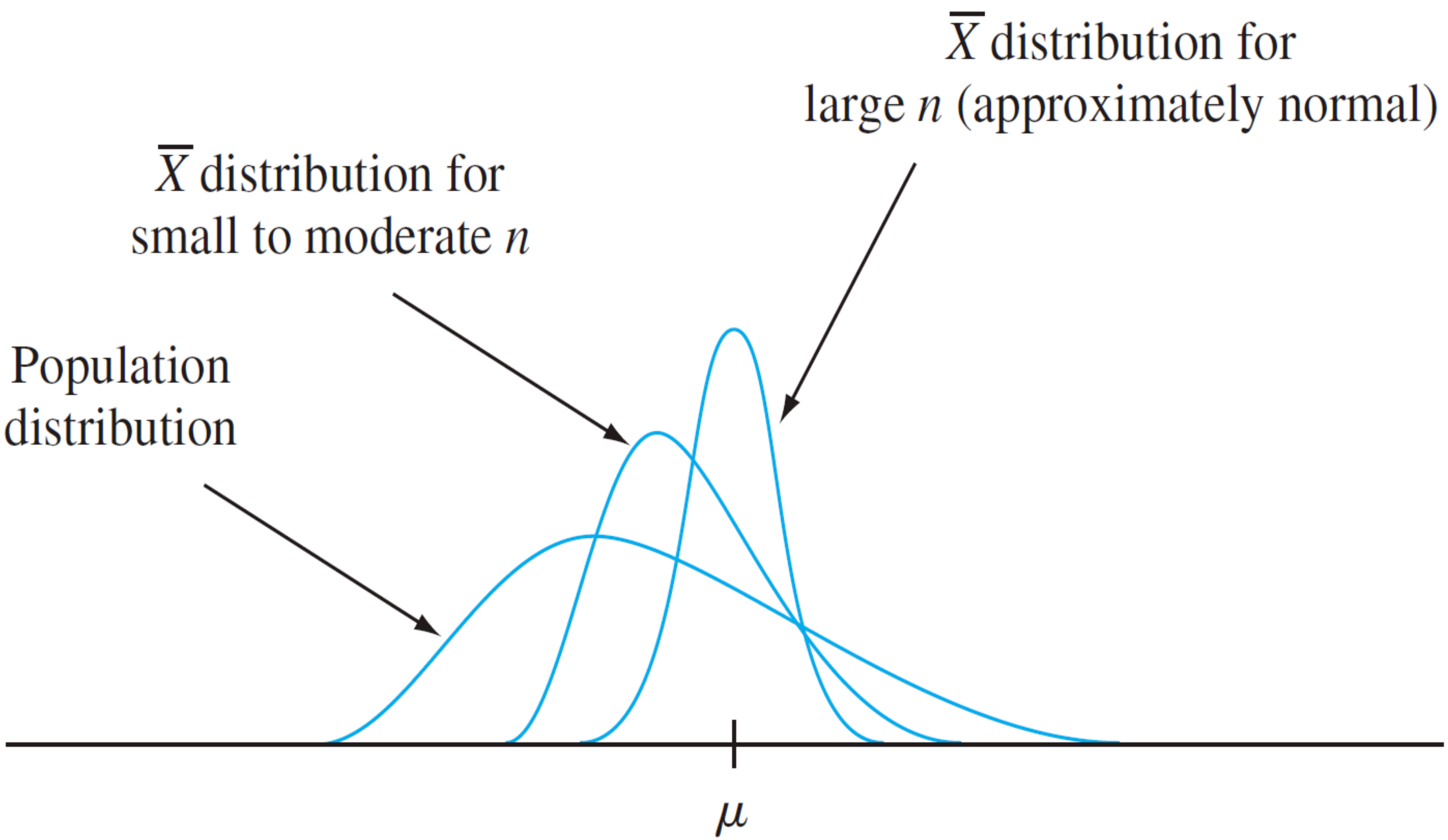
The formal statement of this result is one of the most important theorems in probability: **Central Limit Theorem!**

The Central Limit Theorem

Theorem
The Central Limit Theorem (CLT)

The larger the value of n, the better the approximation! Typical rule of thumb:

The Central Limit Theorem



The Central Limit Theorem

Example: The amount of impurity in a batch of a chemical product is a random variable with mean value 4.0 g and standard deviation 1.5 g. (unknown distribution)

If 50 batches are independently prepared, what is the (approximate) probability that the average amount of impurity in these 50 batches is between 3.5 and 3.8 g?

The Central Limit Theorem

The CLT provides insight into why many random variables have probability distributions that are approximately normal.

For example, the measurement error in a scientific experiment can be thought of as the sum of a number of underlying perturbations and errors of small magnitude.

A practical difficulty in applying the CLT is in knowing when n is sufficiently large. The problem is that the accuracy of the approximation for a particular n depends on the shape of the original underlying distribution being sampled.

Statistical Inference: Confidence Intervals

Confidence Interval for the Mean (SD known)

Let's start with a simple example. Suppose that we have a simple random sample of n measurements from a normal population, and that the population standard deviation is known.

Standardizing the sample mean by first subtracting its expected value and then dividing by its standard deviation yields the standard normal variable

How big does our sample need to be if the underlying population is normally distributed?

Confidence Interval for the Mean (SD known)

Because the area under the standard normal curve between –1.96 and 1.96 is 0.95, we know:

This is equivalent to:

Confidence Interval for the Mean (SD known)

The interval

Is called the 95% confidence interval for the mean.

This interval varies from sample to sample, as the sample mean varies. So, the interval itself is a random interval.

Interpreting a Confidence Interval

"We are 95% confident that the true parameter is in this interval."

What does that mean??

Interpreting a Confidence Interval

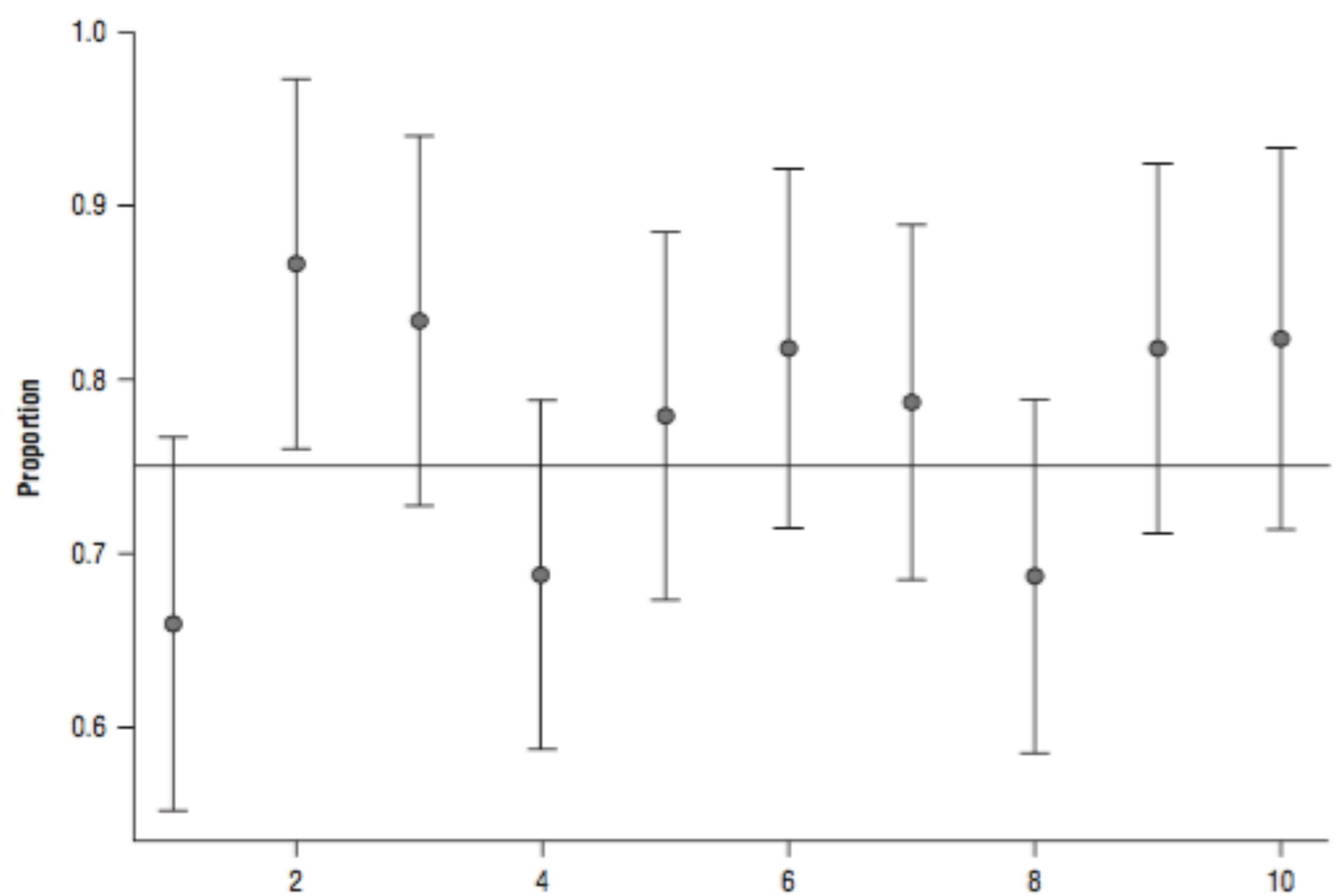
A correct interpretation of “95% confidence” relies on the long-run relative frequency interpretation of probability.

In repeated sampling, 95% of the confidence intervals obtained from all samples will actually contain μ. The other 5% of the intervals will not.

The confidence level is *<u>not a statement about any particular interval</u>* instead it pertains to what would happen if a very large number of like intervals were to be constructed using the same CI formula.

Interpreting a Confidence Interval

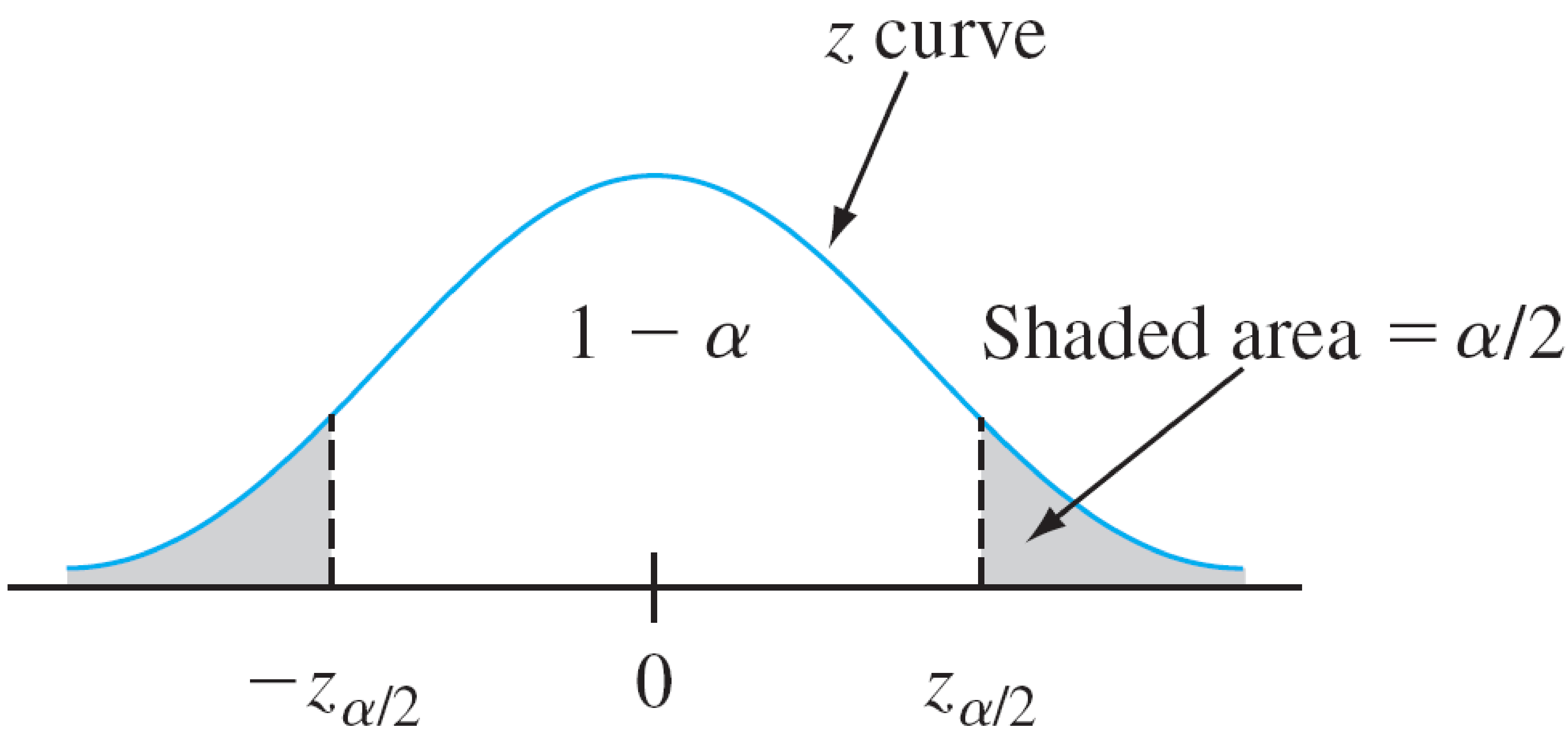
Figure 1: Confidence Interval



Note: Suppose that the true proportion of believers in climate change among French citizens is 0.75, as represented by the horizontal black line near the middle. This figure shows ten 90% confidence intervals used to estimate the proportion of believers in climate change among French citizens. Each circle represents a point estimate, p, calculated from a different sample of n French citizens; for each confidence interval, the length of the vertical line is twice the margin of error, E, for that interval. Notice that the second interval fails to cover the true proportion. For the 90% confidence interval procedure, it is expected that about one in every ten intervals will fail to cover the true proportion.

Other Levels of Confidence

A confidence level of $(1-\alpha)\times 100\%$ is achieved by using $z_{\alpha/2}$ in place of $z_{.025} = 1.96$:



Other Levels of Confidence

So, a **(1 $-\alpha$) x 100% confidence interval** for the mean when the value of σ is known is given by:

Confidence Interval for the Mean (SD Known)

Example: A manufacturing process produces screen protectors for iPhones. A sample of 40 protectors is selected from this process and the width of each protector is measured. The sample mean width is 67.3 mm, and the standard deviation of measurements is 0.1mm.

(a) Calculate a confidence interval for true average protector width using a confidence level of 90%.

(b) What about the 99% confidence interval?

(c) What are the advantages and disadvantages to a wider confidence interval?

Large Sample Confidence Interval for the Mean

A difficulty in using our previous equation for confidence intervals is that it uses the value of σ which will rarely be known. Also, we may want a CI for a mean from some other non-normal distribution.

Large Sample Confidence Interval for the Mean

In this instance, we need to work with the **sample standard deviation** s. With this substitution, we instead work with the standardized random variable:

Confidence Interval for the Mean (SD Unknown)

Previously, there was randomness only in the numerator of Z by virtue of the estimator __________.

In the new standardized variable, both ________ *and* ________ vary in value from one sample to another.

When n is large, the substitution of s for σ adds little extra variability, so nothing needs to change.

When n is smaller, the distribution of this new variable should be wider than the normal to reflect the extra uncertainty. (We talk more about this in a bit.)

Confidence Interval for the Mean (SD Unknown)

<u>Large Sample CI</u>:

If *n* is sufficiently large *(n $\geq$30),* the standardized random variable

has approximately a standard normal distribution. This implies that

is a large-sample confidence interval for with confidence level approximately 100(1 $-\alpha$)%. This formula is valid regardless of the population distribution for sufficiently large *n*.

	n >= 30	n < 30
Underlying normal distribution	**σ known**	**σ known**
	σ unknown	**σ unknown**
Underlying non-normal distribution	**σ known**	**σ known**
	σ unknown	**σ unknown**

Small Sample Interval for the Mean

The CLT cannot be invoked when *n* is small, and we need to do something else when $n < 30$.

When $n < 30$ and the underlying distribution is normal, we have a solution!

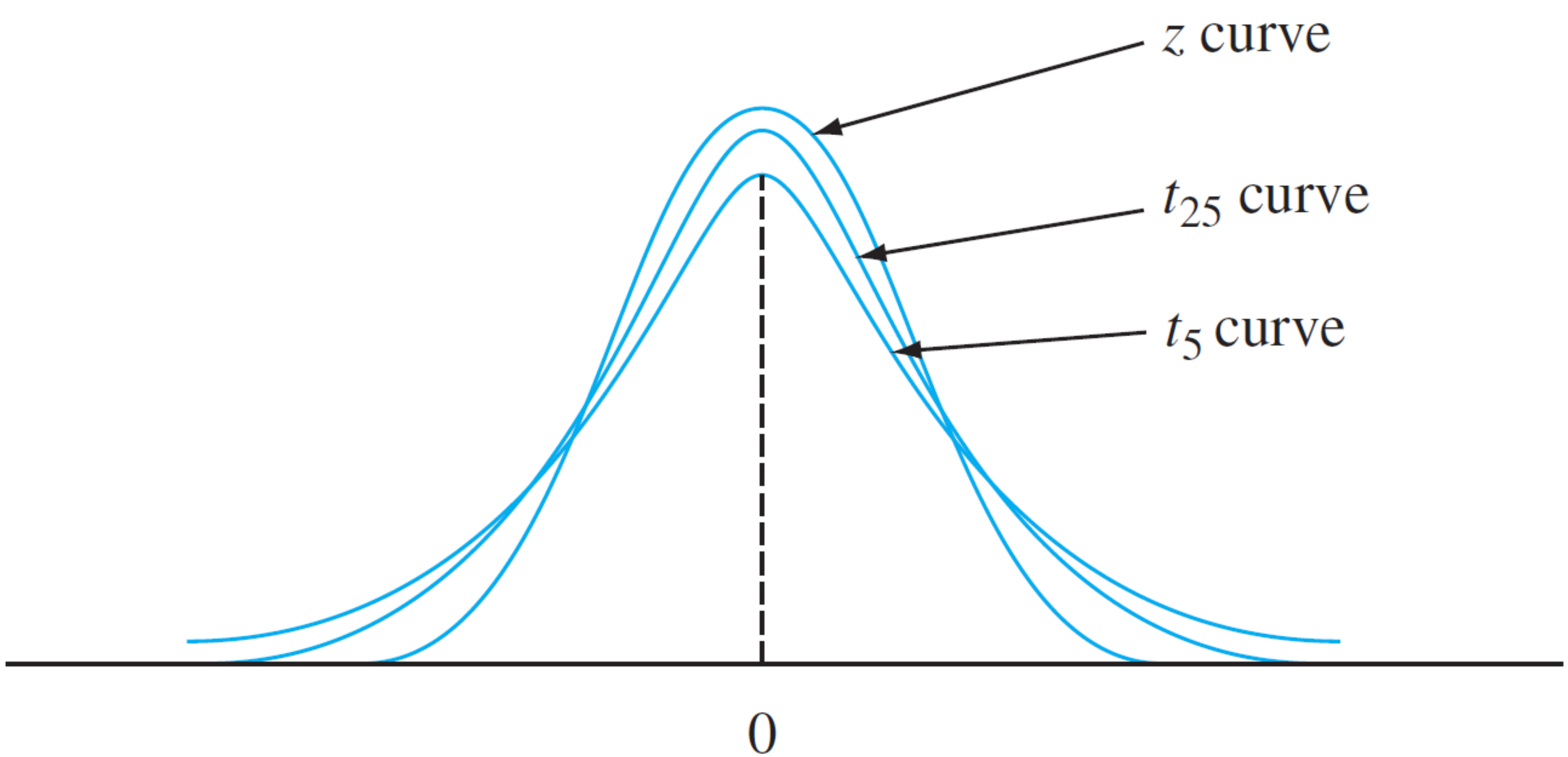
t-Distribution

The results on which large sample inferences are based introduces a new family of probability distributions called *t distributions.*

When ______ is the mean of a random sample of size *n* from a **normal distribution** with mean _________, the random variable

has a probability distribution called a *t* Distribution with n–1 degrees of freedom (df).

t-Distribution



Properties of the t-Distribution

Let t_ν denote the t distribution with ν df.

1. Each t_ν curve is bell-shaped and centered at 0.

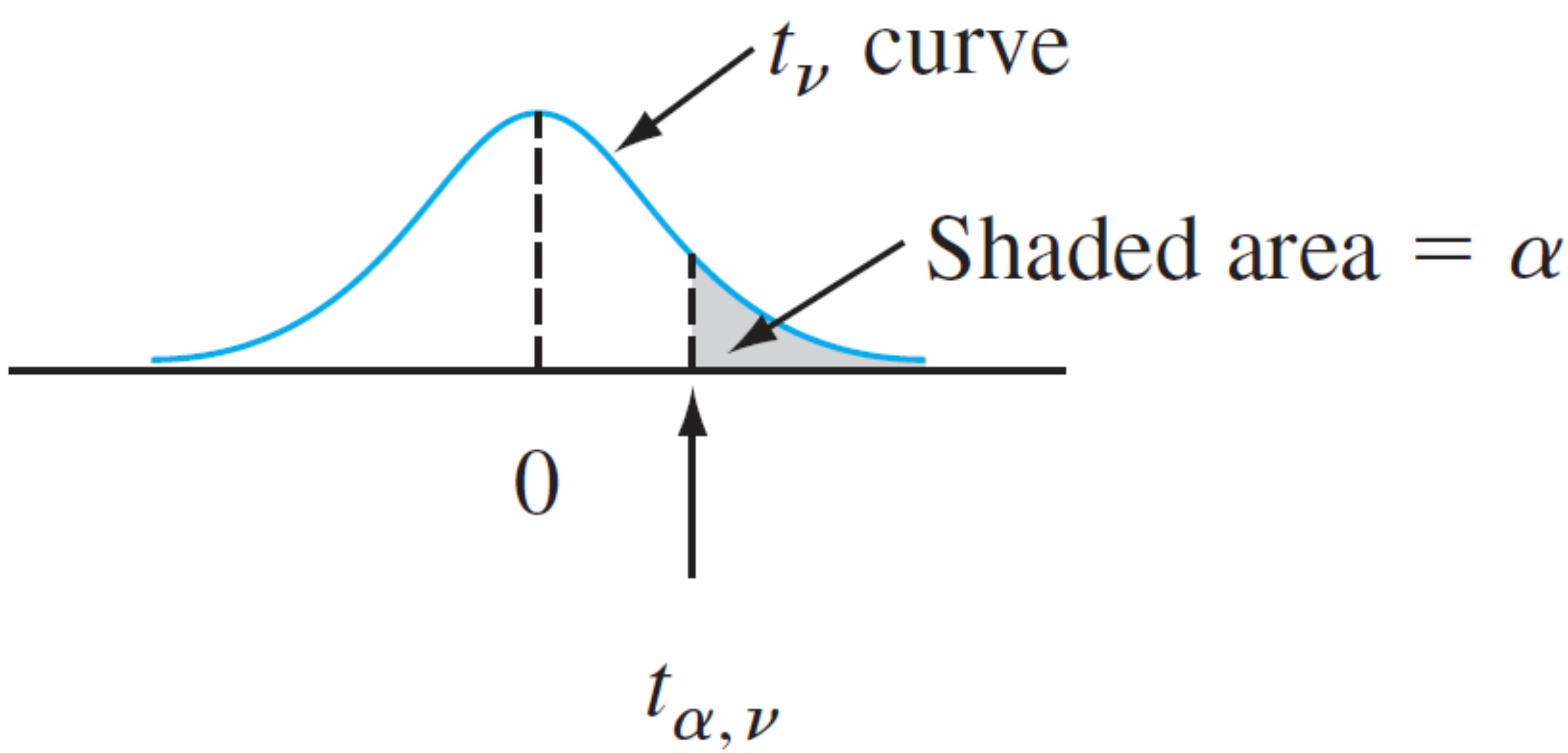
2. Each t_ν curve is more spread out than the standard normal (z) curve.

3. As ν increases, the spread of the corresponding t_ν curve decreases.

4. As ν ______ the sequence of t_ν curves approaches the standard normal curve (so the z curve is the t curve with df = ______)

Properties of the t-Distribution

Let $t_{\alpha,\nu}$ = the number on the measurement axis for which the area under the t curve with ν df to the right of $t_{\alpha,\nu}$ is ; $t_{\alpha,\nu}$ is called a ***t* critical value**.



For example, $t_{.05,6}$ is the t critical value that captures an upper-tail area of .05 under the t curve with 6 df.

Finding t-Values

The probabilities of t curves are found in a similar way as the normal curve.

Example: obtain $t_{.05,15}$

The t-Confidence Interval

Let _____ and ______ be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a $100(1-\alpha)\%$ *t*-confidence interval for the mean μ is

or, more compactly:

The t-Confidence Interval

Example: Suppose that the GPA measurements for 23 students follow a normal distribution. The sample mean is 3.146. The sample standard deviation is 0.308. Calculate a 90% CI for the mean GPA.

Statistical Inference: Hypothesis Tests

Hypotheses

Consider a hypothesis about a random process. For example:

(a) The highest hourly wage within some (large) population of workers is \$50.

(b) The average hourly wage within some (large) population of workers is greater than \$20; or

(c) The chance of getting ‘heads’ on a flip of this coin is 0.5.

Each of these claims is either true or false. But before collecting evidence, we don’t know which! *How could we test these hypotheses, to gain evidence in favor of their truth of falsity?*

Hypotheses

Definition: A (statistical) *hypothesis* is a claim about a population parameter.

Examples: $M1 = M0$ $p=p0$

$s<s0$

Note: There is a distinction to be made between a *research/scientific hypothesis* and a statistical hypothesis. Properly translating a research hypothesis about complex empirical phenomena into a relevant statistical hypothesis is an important skill!

Hypotheses

In any hypothesis-testing problem, there are always two competing hypotheses under consideration:

Definition: The two complementary hypotheses in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis.*

The objective of hypothesis testing is to decide, based on sample information, if the alternative hypotheses is actually supported by the data.

The sample information is summarized by a **test statistic.**

Logic of Hypothesis Testing

Definition: A *test statistic* is a a quantity derived based on sample data and calculated under the null hypothesis. It is used in a decision about whether to reject H_0.

Example:

Suppose im given the null hypothesis, m = m0, Ha = m != m0.
Given this info, a test stat might be, z= standardize which is approximatley N(0,1)

Logic of Hypothesis Testing

Important Question: Is there strong evidence for the alternative? The burden of proof is placed on those who believe in the alternative claim.

The initially favored claim, the null hypothesis H_0, will not be rejected in favor of the alternative hypothesis, H_a or H_1, unless the sample evidence provides a lot of support for the alternative.

The two possible conclusions:

1: reject null in favor of the claim,
2: fail to reject the null

Logic of Hypothesis Testing

(Imperfect) Analogy: Jury in a criminal trial.

A jury in a criminal trial is supposed to presume that the defendant is not guilty (not guilty; that's the "null hypothesis").

Then, the prosecution gathers evidence. If the evidence seems implausible under the assumption of non-guilt, we might reject non-guilt and claim that the defendant is guilty.

Logic of Hypothesis Testing

Why *assume* the null hypothesis?

It represents the status quo
It is reluctant to change
Ethical issues

Innocent before proven guilty

Logic of Hypothesis Testing

What errors are possible?

Logic of Hypothesis Testing

Ex:
H0 p = 1/2
HA p != 1/2

		Your Decision	
		Fail to Reject H_0	Reject H_0
Reality	H_0 is true	No error!	**Type I error (false positive)**
	H_0 is false	**Type II error (false negative)**	no error!

Significance Level

Definition: For a simple null hypothesis, the *significance level* or *size* of the test is the probability of a Type I error.

The significance level is denoted by α:

p(reject H0lH0 is true)

Important! We set the significance level at α and then conduct our test such that the probability of Type II error is minimized (details in Math Stat!).

Conducting a Hypothesis Test

Example: A manufacturing process produces screen protectors for iPhones. The process is supposed to be calibrated to produce screens that are, on average, 67 mm. Engineers collect a random sample of 40 protectors from this process and measure the width of each protector. The sample mean width is 67.3 mm (the population standard deviation of measurements is known to be 0.1mm). The engineers want to decide whether the process is calibrated incorrectly.

(a) What are the null and alternative hypotheses for this test?

H0: m = 67
HA: m != 67

(b) What is a reasonable test statistic?

Use Z test

(c) How could we decide of the test statistic is "sufficiently rare" under the null hypothesis?

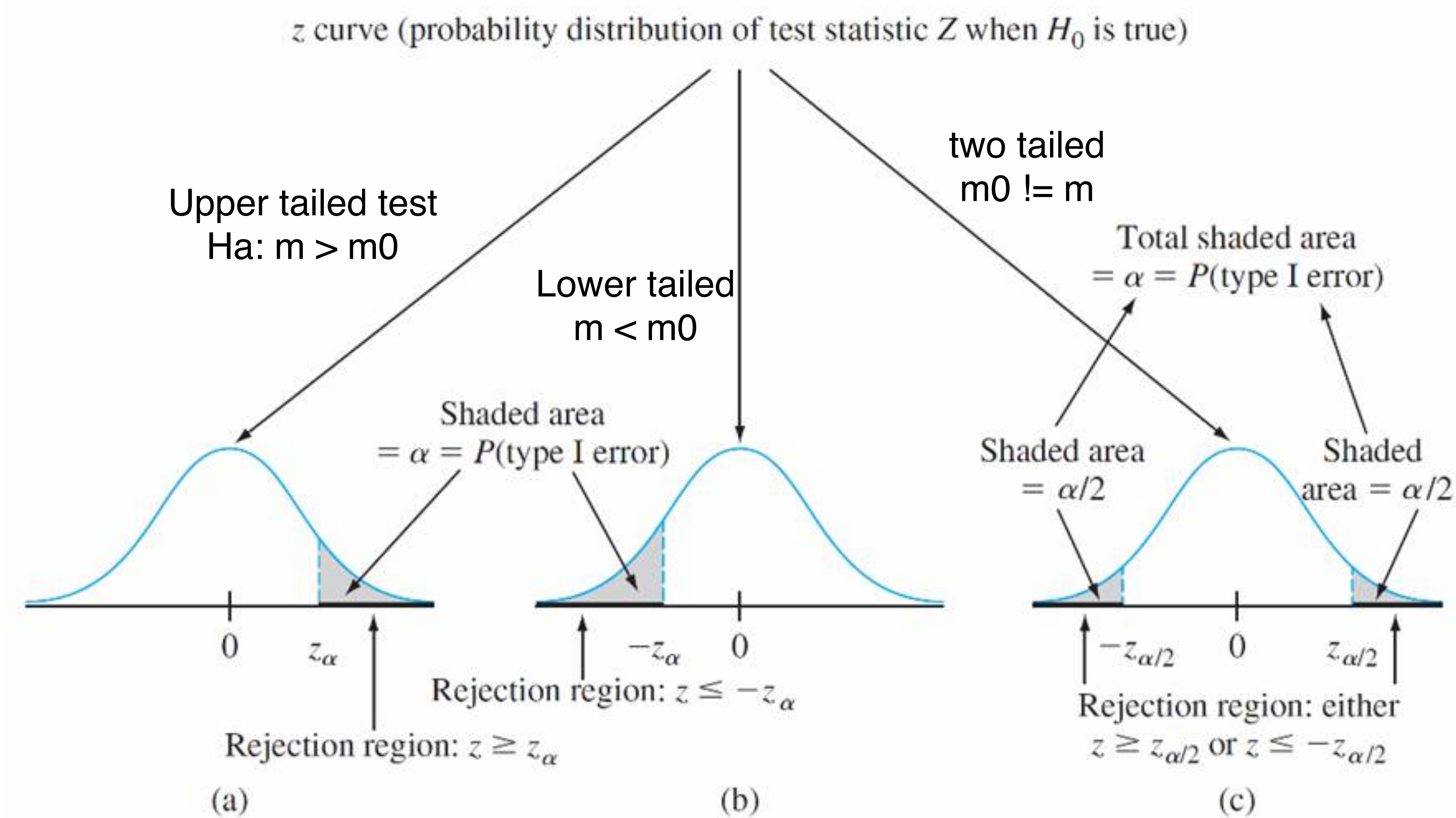
If Z is too high or too low we might reject the null H0

Rejection Regions

How would we know when the test statistic is "sufficiently rare" under the null hypothesis such that we might regard the null as false?

We could define a *rejection region*—a range of values that leads a researcher to *reject* the null hypothesis.

Rejection Regions



Conducting a Hypothesis Test

Example (Continued): A manufacturing process produces screen protectors for iPhones. The process is supposed to be calibrated to produce screens that are, on average, 67 mm. Engineers collect a random sample of 40 protectors from this process and measure the width of each protector. The sample mean width is 67.3 mm (the population standard deviation of measurements is known to be 0.1mm). The engineers want to decide whether the process is calibrated incorrectly.

alpha = .05

(d) State the correct rejection rejoin for the test and decide whether to reject or fail to reject the null hypothesis.

$Z > 1.96$ or $Z < -1.96$

$Z = 18.97 > 1.96$ so reject the null!!!

Test for Population Mean (known variance)

Null hypothesis: H_0 :   m = m0

Test statistic value:

Z = x - M0 / s/sqrt(n) ~ N(0,1)

Alternative Hypothesis   Rejection Region for   level test

Ha 2 tailed: M != M0   Z < Z.025 or Z > Z.025

M > M0

Z < -Z.025

M < M0

Z>Z.025

Practice

Example: An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for more than 5 hours (300 minutes) on a single gallon of regular gasoline. (The leading brand lawnmower engine runs for 300 minutes on 1 gallon of gasoline.)

From his stock of engines, the inventor selects a simple random sample of 50 engines for testing. The engines run for an average of 305 minutes. The true standard deviation σ is known and is equal to 30 minutes, and the run times of the engines are normally distributed.

Test hypothesis that the mean run time is more than 300 minutes. Use a 0.01 level of significance.

Ha: M > 300
Ho: M = 300

Z = x - m / s/sqrt(n)

= 305-300/30/sqrt(50)

= 1.18

Z.01 = 2.33 > 1.18
Fail to reject!!!

Testing Means for a Large Sample

When the sample size is large, the *z* tests are easily modified to yield valid test procedures without requiring either a normal population distribution or known standard deviation.

Earlier, we used the key result to justify large-sample confidence intervals:

A large *n (>30)* implies that the standardized variable

has *approximately* a standard normal distribution.

Testing Means for a Small Sample

When the sample size is small and the population is normal, we can use a **t-test**.

Null hypothesis: H_0 :

Test statistic value:

Alternative Hypothesis   Rejection Region for α level test

P-Values

The p-value measures the “extremeness” of the test statistic.

Definition: A *p-value* is the probability**, under the null hypothesis,** that we would get a test statistic *at least as extreme as the one we calculated.*

So, the smaller the p-value, the more evidence there is in the sample data against the null hypothesis (so the story goes…).

So what constitutes “sufficiently small” and “extreme enough” to make a decision about the null hypothesis?

P-Values

Select a significance level (as before, the desired *type I error probability*), then the p-value defines the rejection region. The decision rule is:

if p <= alpha, reject the null
if p > alpha, fail to reject the null

Thus if the *p*-value exceeds the chosen significance level, the null hypothesis cannot be rejected at that level.

Note, the *p*-value can be thought of as the <u>smallest significance level</u> at which H_0 can be rejected.

P-Values

The p-value measures the “extremeness” of the test statistic.

A very small p-value does not imply a large effect size

Note:

(a) This probability is calculated assuming that the null hypothesis is true.

(b) Beware: The *p*-value is not the probability that H_0 is true, nor is it an error probability!

(c) The *p*-value is between 0 and 1.

P-Values for Z Tests

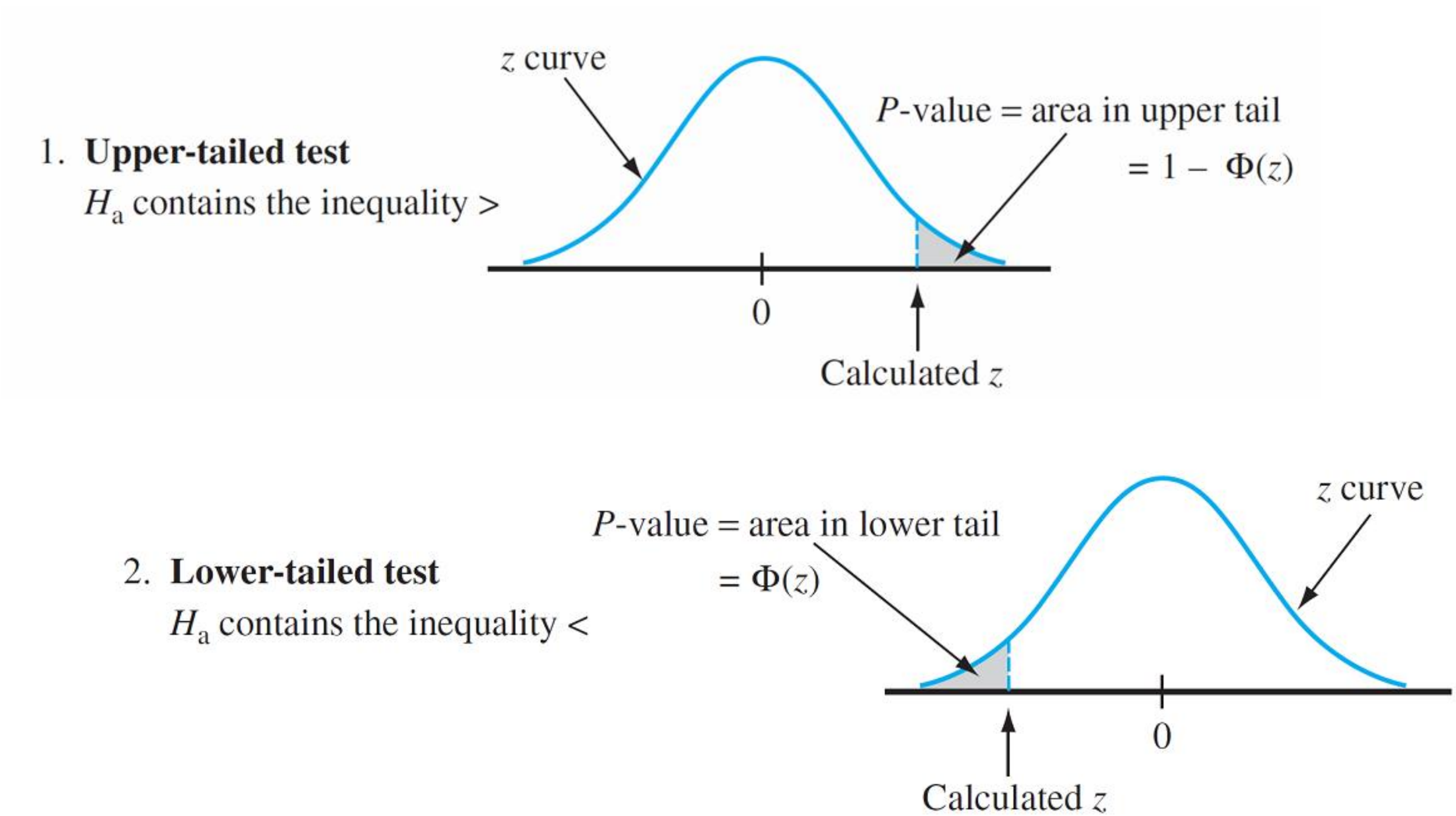
The calculation of the *p*-value depends on whether the test is upper-, lower-, or two-tailed.

phi(z) = pnorm(z)

p = pvalue =
1 - phi(z) - upper tailed
phi(z) - lower
2(1-phi(z)) - two-tailed

Each of these is the probability of getting a value at least as extreme as what was obtained (assuming H_0 true).

P-Values for Z Tests



P-Values for Z Tests

3. **Two-tailed test**
   H_a contains the inequality $\neq$