STA261 Lecture 9 — 2017-08-09

Neil Montgomery

Last edited: 2017-08-09 19:05

interval estimation so far

An interval estimate (i.e. "a $(1-\alpha)\cdot 100\%$ confidence interval") for θ is a pair of random variables L, U such that

$$P(L < \theta < U) = 1 - \alpha$$

One technique to getting a confidence interval is to find a *pivot*.

That being said, the vast majority of all confidence intervals encountered in actual practice are of the 95% variety and look like this:

$$\hat{\theta} \pm \text{"2"} SD(\hat{\theta})$$

where "2" is a number close to 2.

“margin of error” and “standard error”

The number after the $\pm$ in such an interval is often called the “margin of error”.

The standard deviation of $\hat{\theta}$ often contains other parameters or even other estimators for those parameters. In the $N(\mu, \sigma^2)$ case, $SD(\hat{\mu})$ is

$$\frac{S}{\sqrt{n}}$$

As an abbreviation this is called the “standard error”.

large sample confidence intervals

Recall from the likelihood theory:

$$\sqrt{\mathcal{I}(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1)$$

large sample confidence intervals

Recall from the likelihood theory:

$$\sqrt{\mathcal{I}(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1)$$

This result is perhaps ultimately the source of most confidence intervals reported from actual data analyses.

large sample confidence intervals

Recall from the likelihood theory:

$$\sqrt{\mathcal{I}(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1)$$

This result is perhaps ultimately the source of most confidence intervals reported from actual data analyses.

An extension of the result is:

$$\sqrt{\widehat{\mathcal{I}(\theta)}}(\hat{\theta}_n - \theta) \Rightarrow N(0,1)$$

where $\widehat{\mathcal{I}(\theta)}$ has (possibly) other parameters replaced with estimates from the data.

“Wald intervals”

For large samples the (approximate) interval is:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{\mathcal{I}(\theta)}}$$

Example 9.0: $N(\mu, \sigma^2)$ (the “base example” from which all others flow, by way of analogy)

"Wald intervals"

For large samples the (approximate) interval is:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{\mathcal{I}(\theta)}}$$

Example 9.0: $N(\mu, \sigma^2)$ (the "base example" from which all others flow, by way of analogy)

Example 9.1: Poisson(λ)

"Wald intervals"

For large samples the (approximate) interval is:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{\mathcal{I}(\theta)}}$$

Example 9.0: $N(\mu, \sigma^2)$ (the "base example" from which all others flow, by way of analogy)

Example 9.1: Poisson(λ)

Example 9.2: "Censored" exponential data (from Lecture 5 and 2017-07-26 tutorial)

“Wald intervals”

For large samples the (approximate) interval is:

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\widehat{\mathcal{I}(\theta)}}$$

Example 9.0: $N(\mu, \sigma^2)$ (the “base example” from which all others flow, by way of analogy)

Example 9.1: Poisson(λ)

Example 9.2: “Censored” exponential data (from Lecture 5 and 2017-07-26 tutorial)

Note of caution: convergence to $N(0,1)$ not always that fast - very large samples might be needed.

(classical) hypothesis testing (note: hypothesis testing is wierd)

Two motivating examples

1. You have n people taking the current standard medication (the "control" group) and m people taking a new medication (the "treatment" group). The results (some sort of blood test) are modeled as:

   $$X_1, \ldots, X_n \text{ i.i.d. } N(\mu_X, \sigma^2) \qquad Y_1, \ldots, Y_m \text{ i.i.d. } N(\mu_Y, \sigma^2)$$

   What is the mathematical statement that embodies "no difference"?

Two motivating examples

1. You have n people taking the current standard medication (the "control" group) and m people taking a new medication (the "treatment" group). The results (some sort of blood test) are modeled as:

$$X_1, \ldots, X_n \text{ i.i.d. } N(\mu_X, \sigma^2) \qquad Y_1, \ldots, Y_m \text{ i.i.d. } N(\mu_Y, \sigma^2)$$

   What is the mathematical statement that embodies "no difference"?

2. You have a set of constants $x_1, \ldots, x_n$ and have observed a sequence of measurements from the following model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

   with ε_i i.i.d. $N(0, \sigma^2)$. What is the mathematical statement that embodies "no relationship"?

hypothesis testing setup

A *hypothesis* is a statement about a population parameter.

The goal of (classical) hypothesis testing is to choose, using data, between one of two hypotheses that contain no parameter values in common.

hypothesis testing setup

A *hypothesis* is a statement about a population parameter.

The goal of (classical) hypothesis testing is to choose, using data, between one of two hypotheses that contain no parameter values in common.

The two hypotheses are called the *null* and *alternative* hypothesis, denoted by H_0 and H_1 or H_a respectively.

hypothesis testing setup

A *hypothesis* is a statement about a population parameter.

The goal of (classical) hypothesis testing is to choose, using data, between one of two hypotheses that contain no parameter values in common.

The two hypotheses are called the *null* and *alternative* hypothesis, denoted by H_0 and H_1 or H_a respectively.

Typically (i.e. not mathematically) the null hypothesis embodies "no difference", "no relationship", etc.

hypothesis testing setup

A *hypothesis* is a statement about a population parameter.

The goal of (classical) hypothesis testing is to choose, using data, between one of two hypotheses that contain no parameter values in common.

The two hypotheses are called the *null* and *alternative* hypothesis, denoted by H_0 and H_1 or H_a respectively.

Typically (i.e. not mathematically) the null hypothesis embodies "no difference", "no relationship", etc.

In my scientific opinion, the alternative hypothesis should almost always just be the complement of the null hypothesis (but there is no mathematical requirement for this.)

examples | the possible decisions

Examples (from the two sample and regression motivators.)

examples | the possible decisions

Examples (from the two sample and regression motivators.)

A (classical) hypothesis test will produce a decision rule that maps data to one of the following decisions:

1. reject H_0

examples | the possible decisions

Examples (from the two sample and regression motivators.)

A (classical) hypothesis test will produce a decision rule that maps data to one of the following decisions:

1. reject H_0
2. not reject H_0

examples | the possible decisions

Examples (from the two sample and regression motivators.)

A (classical) hypothesis test will produce a decision rule that maps data to one of the following decisions:

1. reject H_0
2. not reject H_0

examples | the possible decisions

Examples (from the two sample and regression motivators.)

A (classical) hypothesis test will produce a decision rule that maps data to one of the following decisions:

1. reject H_0
2. not reject H_0

Hypothesis testing is rife with misleading and emotional language. Be careful!!

types of errors and their probabilities

To reject H_0 when it is true is to make a *Type I Error*.

types of errors and their probabilities

To reject H_0 when it is true is to make a *Type I Error*.

To not reject H_0 when it is true is to make a *Type I Error*.

types of errors and their probabilities

To reject H_0 when it is true is to make a *Type I Error*.

To not reject H_0 when it is not true is to make a *Type I Error*.

The probability that a test (i.e. the decision rule) will result in a Type I Error is called α, and has synonyms “significance level” and “size”.

types of errors and their probabilities

To reject H_0 when it is true is to make a *Type I Error*.

To not reject H_0 when it is not true is to make a *Type I Error*.

The probability that a test (i.e. the decision rule) will result in a Type I Error is called α, and has synonyms "significance level" and "size".

The probability that a test will result in a Type II Error is called β.

types of errors and their probabilities

To reject H_0 when it is true is to make a *Type I Error*.

To not reject H_0 when it is not true is to make a *Type I Error*.

The probability that a test (i.e. the decision rule) will result in a Type I Error is called α, and has synonyms "significance level" and "size".

The probability that a test will result in a Type II Error is called β.

The probability $1 - \beta$ is called the *power* of the test.

other definitions

other definitions

The function of the data that is used in the test is called the *test statistic*.

other definitions

The function of the data that is used in the test is called the *test statistic*.

The distribution of the test statistic when H_0 is true is called the *null distribution*.

other definitions

The function of the data that is used in the test is called the *test statistic*.

The distribution of the test statistic when H_0 is true is called the *null distribution*.

The data values that result in H_0 being rejected is called the *rejection region* or *critical region*.

illustration of the ideas.

Example 9.3: $X_1, \ldots, X_{25}$ are i.i.d. $N(\mu, 1)$

$H_0 : \mu = 1$ and $H_1 : \mu = -1$

illustration of the ideas.

Example 9.3: $X_1, \ldots, X_{25}$ are i.i.d. $N(\mu, 1)$

$H_0 : \mu = 1$ and $H_1 : \mu = -1$

The data: 1, 0.8, -0.4, 0.4, 1.3, 1.4, -0.2, 0.6, -0.6, 0.7, 2.1, 1.8, 0.8, 2, 1.7, 1.1, 0, 0.8, 1.9, 1.5, 0.4, -1.2, 0.3, -1.1, -0.3

illustration of the ideas.

Example 9.3: $X_1, \ldots, X_{25}$ are i.i.d. $N(\mu, 1)$

$H_0 : \mu = 1$ and $H_1 : \mu = -1$

The data: 1, 0.8, -0.4, 0.4, 1.3, 1.4, -0.2, 0.6, -0.6, 0.7, 2.1, 1.8, 0.8, 2, 1.7, 1.1, 0, 0.8, 1.9, 1.5, 0.4, -1.2, 0.3, -1.1, -0.3

Some possible decision rules:

1. reject H_0 if the majority of observations are less than 0.

illustration of the ideas.

Example 9.3: $X_1, \ldots, X_{25}$ are i.i.d. $N(\mu, 1)$

$H_0 : \mu = 1$ and $H_1 : \mu = -1$

The data: 1, 0.8, -0.4, 0.4, 1.3, 1.4, -0.2, 0.6, -0.6, 0.7, 2.1, 1.8, 0.8, 2, 1.7, 1.1, 0, 0.8, 1.9, 1.5, 0.4, -1.2, 0.3, -1.1, -0.3

Some possible decision rules:

1. reject H_0 if the majority of observations are less than 0.
2. reject H_0 if $\overline{X} > 0$

illustration of the ideas.

Example 9.3: $X_1, \ldots, X_{25}$ are i.i.d. $N(\mu, 1)$

$H_0 : \mu = 1$ and $H_1 : \mu = -1$

The data: 1, 0.8, -0.4, 0.4, 1.3, 1.4, -0.2, 0.6, -0.6, 0.7, 2.1, 1.8, 0.8, 2, 1.7, 1.1, 0, 0.8, 1.9, 1.5, 0.4, -1.2, 0.3, -1.1, -0.3

Some possible decision rules:

1. reject H_0 if the majority of observations are less than 0.
2. reject H_0 if $\overline{X} > 0$
3. reject H_0 if $\frac{L(1)}{L(-1)} > 1$

likelihood ratio | Neyman-Pearson

Theorem 9.4: Suppose $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ (each contain one value only.) The test of size α that rejects H_0 when:

$$\frac{L(\theta_0)}{L(\theta_1)} < c$$

for some constant c is at least as powerful as any test of size α.

(proof to be given next class)

principal example

Example 9.5: $N(\mu, \sigma^2)$ (σ^2 known) with $H_0 : \mu = \mu_0$ and $H_1 : \mu = \mu_1$ with $mu_1 > \mu_0$.