Introduction to the Theory of Statistics Part 2

PM522b

Meredith Franklin

Division of Biostatistics, University of Southern California

Slides 5, 2018

Topics covered

1. Hypothesis Testing
2. Confidence Intervals (interval estimation)

Hypothesis Testing

- We now turn to hypothesis testing whereby we form a statement about a parameter θ and then perform a statistical test to determine the correctness of the statement.
- Hypothesis testing is similar to the scientific method: a scientist formulates a theory and then tests this theory against observation.
- In statistics we pose a theory concerning one or more population parameters (i.e. that they equal specified values), we then sample the population and compare our observations with our posed theory. If the observations disagree with the theory then we reject it.

Hypothesis Testing

Null and Alternative Hypotheses

- In parametric inference, a statistical hypothesis is a statement concerning an unknown parameter for the population distribution $f(x|\theta), x \in \mathbb{R}$ and $\theta \in \Theta$
- The statistical hypothesis is a statement about θ and the testing aims to prove its correctness.
- The hypothesis specifies that θ belongs to some subset of Θ. We define $\Theta_0 \subseteq \Theta$ and $\Theta_1 \subseteq \Theta$ with $\Theta_0 \cap \Theta_1 = 0$.
- The statement that $\theta \in \Theta_0$ is denoted by H_0 and is a statistical hypothesis that invalidates (nullifies) the statement under investigation. This is the null hypothesis.
- The statement that $\theta \in \Theta_1$ is denoted by H_1 and is the statistical hypothesis under investigation. This is the alternative hypothesis.
- In practice, after observing a sample it must be decided whether to accept (fail to reject) H_0 or to reject H_0 and decide H_1 is true.
- The alternative hypothesis H_1 is usually taken to be the negation of H_0.

Hypothesis Testing

Simple and Composite Hypotheses

- A hypothesis $H_i : \theta \in \Theta_i, i = 0, 1$ is called *simple* if the subset Θ_i contains only one element, $\Theta_i = \{\theta_i\}$
- Under a simple hypothesis, $H_i : \theta = \theta_i$, the population distribution $f(x|\theta_i), x \in \mathbb{R}$ is completely specified
- A hypothesis $H_i : \theta \in \Theta_i, i = 0, 1$ is called *composite* in Θ_i contains more than one element
- Under a composite hypothesis, $H_i : \theta = \theta_i$, the population distribution belongs to a family of distribution $f(x|\theta_i), x \in \mathbb{R}, \theta \in \Theta_i, i = 0, 1$
- There are two forms of composite hypotheses:
  - *one-sided* where they take the form $H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$ for the null hypothesis and subsequently $H_1 : \theta > \theta_1$ or $H_1 : \theta < \theta_1$ for the alternative hypothesis
  - *two-sided* where they take the form $H_0 : \theta = \theta_0$ for the null hypothesis and subsequently $H_1 : \theta \neq \theta_0$ for the alternative hypothesis

Hypothesis Testing

Critical and Acceptance Regions

- A statistical test of the null hypothesis $H_0 : \theta \in \Theta_0$ is a procedure by which, using the observed values $x_1, ..., x_n$ of a random sample $X_1, ..., X_n$, we come to the decision to reject H_0 and accepting H_1 OR accepting H_0 (not rejecting H_0).
- The subset of the sample space for which H_0 will be rejected is called the *rejection* or *critical region*, R.
- The compliment to the critical region is the *acceptance region*, A
- We need a test statistic $T(X_1, ..., X_n)$ that partitions (or maps) $x_1, ..., x_n$ into these two subset regions R and A.
- For example, a simple test could be that if $T(X_1, ..., X_n) = 1$ then we reject H_0, thus $R = \{(x_1, ..., x_n) : T(x_1, ..., x_n) = 1\}$ is our critical region.
- Another example, we could define a test where the critical region is $R = \{(x_1, ..., x_n) : \bar{x} > 0\}$.

Hypothesis Testing

In Class Example: Suppose a mayoral candidate claims she will get more than 50% of the votes in an election, and thereby be the winner. We do not believe this claim, so we would like to test the candidate's claim as a hypothesis test. Set up the elements of a statistical test (hypothesis test):

- null hypothesis
- alternative hypothesis
- test statistic
- rejection/critical region

Hypothesis Testing

Type I and Type II errors

- A statistical test of the null hypothesis $H_0 : \theta \in \Theta_0$ against an alternative hypothesis $H_1 : \theta \in \Theta_1$ leads to either a correct decision or one of the following errors:
  - Type I error = rejecting H_0 when it is true
  - Type II error = accepting H_0 when it is false

	H_0 is accepted	H_0 is rejected
H_0 is true	correct decision	type I error
H_0 is false	type II error	correct decision

Hypothesis Testing

Type I and Type II errors

The probability of rejecting the null hypothesis with the parameter θ restricted on the subsets Θ_0 and Θ_1 of the parameter space Θ can be expressed as the following:

Suppose the null hypothesis is true, $\theta \in \Theta_0$. The probability we will make an error in our decision occurs when our test statistic $T(X)$ falls in the rejection region: $P_\theta(X \in R) = \alpha$. This is the Type I error. Using this same logic, the probability that our test statistic $T(X)$ falls in the acceptance region is

$P_\theta(X \in A) = 1 - P_\theta(X \in R)$

Suppose the alternative hypothesis is true, $\theta \in \Theta_1$. The probability we will make an error in our decision occurs when our test statistic $T(X)$ falls in the acceptance region: $P_\theta(X \in A)$. This is the Type II error. If we think about the probability that we reject the null hypothesis supposing the alternative hypothesis is true, we obtain the power. That is, for $\theta \in \Theta_1$, $P_\theta(X \in R) = \beta(\theta)$.

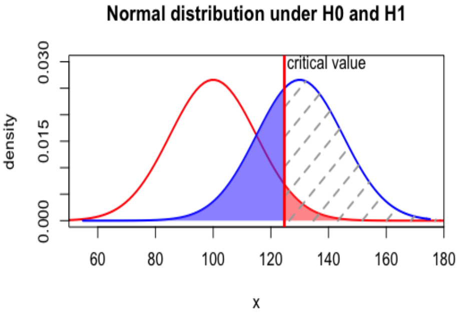
Hypothesis Testing

Power of a test (power function)

- $\beta(\theta) = P_\theta[(X_1, X_2, ..., X_n) \in R], \theta \in \Theta_1$ is the probability of rejecting H_0 given that H_1 is correct, and this is the correct decision of rejecting H_0
- The function $\beta(\theta)$ is called the **power** function of the the test and its value at a specific point $\theta = \theta_1 \in \Theta_1$ is the **power of the test** at θ_1
- The ideal power function is 0 for all $\theta \in \Theta_0$ and 1 for all $\theta \in \Theta_1$
- In a statistical test, it is desirable to keep the probabilities of Type I and Type II errors small. In searching for a good test, commonly the tests are restricted to control the Type I error probability at a specified level. Then within this class of tests, we search for those that have the smallest Type II error probability
- In controlling Type I error probabilities we have:
  - The **size** of a test, defined as $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$ for $0 \leq \alpha \leq 1$
  - The **level** of a test, defined as $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$ for $0 \leq \alpha \leq 1$

Hypothesis Testing

Example where we have expected value μ_0 under H_0 (100), expected value μ_1 under H_1 (130)



Hypothesis Testing

Significance Level and Most Powerful Test

- In testing a statistical hypothesis, a significance level α with $0 \leq \alpha \leq 1$ is taken among the tests satisfying

$$\alpha(\theta) \leq \alpha \text{ for all } \theta \in \Theta_0$$

- The test that minimizes the probability of Type II error for all $\theta \in \Theta_1$, or equivalently that maximizes the power for all $\theta \in \Theta_1$ is chosen.
- If the set Θ_1 contains one point, $\Theta_1 = \{\theta_1\}$, then the test is called the *most powerful test*
- If the set Θ_1 contains more than one point then the test is called the *uniformly most powerful test* (UMP)
- Usually significance level is chosen to be $\alpha \leq 0.05$, but a more conservative significance level is $\alpha \leq 0.01$.

Hypothesis Testing

p-Values

- It has become standard practice to report the **size** of the test, α used in the decision to reject or accept H_0 as it carries important information:
  - If α is small, the decision to reject H_0 is convincing (giving evidence that H_1 is true)
  - If α is large (usually larger than 0.05), the decision to reject H_0 is not very convincing because the hypothesis test has a large probability of incorrectly making that decision
- For a sample $x_1, ..., x_n$, a p-value is a test statistic satisfying $0 \leq p(x) \leq 1$, a p-value $p(X)$ is valid if

$$P(p(X) \leq \alpha) \leq \alpha$$

- The general approach for defining a valid p-value is to determine a test statistic $T(X)$ such that large values of T(x) (the observed value of the test statistic) give evidence that H_1 is true

$$p(x) = \sum_{\theta \in \Theta_0} P(T(X) \geq T(x))$$

Likelihood Ratio Test

The likelihood ratio test (LRT) is a very general method of deriving tests of hypothesis. The procedure works for both simple or composite hypotheses.

Likelihood Ratio Test

This method of hypothesis testing is related to maximum likelihood estimators.

- The generalized LRT is formulated as the ratio of the maximum probability of the observed sample being computed over the parameters in the null hypothesis, H_0 to the maximum probability of the observed sample over *all* possible parameters in Θ.

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta|x)}{\sup_{\Theta} L(\theta|x)}$$

  $\lambda(x)$ is the likelihood ratio test statistic.

- Consider testing the null $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1$. In terms of a simple hypothesis, $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$. The LRT statistic for this test is:

$$\lambda(x) = \frac{L(\theta_0|x)}{L(\theta_1|x)}$$

- The test for these statistics is defined by

$$R = \{x : \lambda(x) \leq c\} \text{ for } 0 \leq c \leq 1$$

Likelihood Ratio Test

Unrestricted and restricted maximization

- $\hat{\theta}$ is the MLE of θ, obtained by doing a maximization over all possible parameters Θ
- $\hat{\theta}_0$ can also be an MLE of θ, but obtained by doing a maximization over the restricted parameter space Θ_0
- That means $\hat{\theta}_0 = \hat{\theta}_0(x)$ is the value of the parameter $\theta \in \Theta_0$ that maximizes the likelihood $L(\theta|x)$. In this case, the LRT is

$$\lambda(x) = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}$$
$$= \frac{\sup\limits_{\theta \in \Theta_0} L(\theta|x_1, ..., x_n)}{\sup\limits_{\theta \in \Theta} L(\theta|x_1, ..., x_n)}$$

Likelihood Ratio Test

- The critical region for the LRT statistic is defined by:

$$R = \{x_1, ..., x_n : \lambda(x) \leq c\}$$

where c is a constant between 0 and 1 and will be for a given significance level α when it is chosen to satisfy:

$$\sup_{\theta \in \Theta_0} \{P[\lambda(x) \leq c_\alpha | \theta \in \Theta_0]\} = \alpha$$

Likelihood Ratio Test

Example: LRT of normal distribution

Testing the mean of $\mathrm{N}(\mu, \sigma^2)$ where σ^2 is known, namely $\mathrm{H}_0 : \mu = \mu_0$ against $\mathrm{H}_1 : \mu \neq \mu_0$. μ_0 is a number fixed by the experimenter before doing the experiment.

$$L(\mu|x_1, ..., x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2]$$

$$\sup_{\mu\in\{\mu_0\}} L(\mu|x_1, ..., x_n) = L(\mu_0|x_1, ..., x_n)$$

$$\sup_{\mu\in\Theta} L(\mu|x_1, ..., x_n) = L(\hat{\mu}|x_1, ..., x_n)$$

where $\hat{\mu} = \bar{x}$ is the MLE of μ

$$\lambda(x) = \frac{L(\mu_0|x_1, ..., x_n)}{L(\hat{\mu}|x_1, ..., x_n)} = \frac{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2]}{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2]}$$

Likelihood Ratio Test

Example: LRT of normal distribution con't

Noting that

$$\sum_{i=1}^{n}(x_i - \mu_0)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2$$

the LRT statistic becomes

$$\frac{n(\bar{x} - \theta_0)^2}{\sigma^2} = \left(\frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}}\right)^2$$

Because $\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}$ is a standard normal random variable, the above $\Lambda(x)$ is the square of two standard normals which is a χ^2_1 random variable.

Likelihood Ratio Test

Example: LRT of normal distribution con't

$$\left(\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}\right)^2 \text{ and } \left|\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}\right|$$

Both are test statistics, with the first being the square of the critical value for the second. The critical region in terms of the second is:

$$R = \{x_1, ..., x_n : \left|\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}\right| \geq c\}$$

where c is a constant for a given significance level α, determined by

$$P(\left|\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}\right| \geq c|\mu = \mu_0) = \alpha$$

Since $\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}$ is a standard normal random variable we have a two-sided z-test. The test is to reject H_0 if $\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}} \geq z_{\alpha/2}$.

Neyman-Pearson Lemma

This Lemma allows us to find the test of a given size α with the largest power (it is a most powerful test) and formulated around the simple hypothesis testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ where $\theta_0 \neq \theta_1$.
Basically the Neyman-Pearson lemma tells us that the best test for a simple hypothesis is a likelihood ratio test.

Neyman-Pearson Lemma

Neyman-Pearson Lemma

Let $X_1, .., X_n$ be a random sample from a distribution with parameter θ, where $\theta \in \Theta = \{\theta_0, \theta_1\}$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ where $\theta_0 \neq \theta_1$; the pdf or pmf corresponding to $\theta_i, i = 0, 1$ is $f(x|\theta_i)$ and corresponding likelihood function is $L(\theta_i|x)$.
If there exists a test at significance level α such that for some positive constant k,

$$\frac{L(\theta_1|x)}{L(\theta_0|x)} \geq k \text{ for each } x \in R \text{ (inside the critical region)}$$
$$\frac{L(\theta_1|x)}{L(\theta_0|x)} \leq k \text{ for each } x \notin R \text{ (outside the critical region)}$$

then R is the most powerful critical region of size α, and this is the most powerful test for testing the null against alternative hypothesis, $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Neyman-Pearson Lemma

The Neyman-Pearson lemma gives us the most powerful test of size α for $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.
Thus, the test based on the critical region $R = \{x : \lambda(x) \leq k\}$ has the largest power (smallest Type II error) of all tests with significance level α.
Among all tests with a given probability of Type I error, the likelihood ratio test minimizes the probability of a Type II error.

Proof Neyman-Pearson Lemma

For a continuous random variable, let R be the critical region of size α and A be another region of size α, both fitting the conditions of the Neyman-Pearson Lemma. Also, let $\int \cdots \int L(\theta|x_1, ..., x_n)dx_1, ..., dx_n$ be represented by $\int L(\theta)$ (integrated over a region).
Proof not required, shown on pp. 388-389 of CB.

Neyman-Pearson Lemma

Example: let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$ where μ is unknown and σ^2 is known. For the hypothesis $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$ where $\mu_1 > \mu_0$ find the most powerful test.

$$\frac{L(\theta_1|x)}{L(\theta_0|x)} = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu_1)^2]}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu_0)^2]}$$

$$= \frac{\exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu_1)^2}{\exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu_0)^2}$$

$$= \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n}((x_i - \mu_1)^2 - (x_i - \mu_0)^2)$$

$$= \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i^2 - 2\mu_1 x_i + \mu_1^2) - \sum_{i=1}^{n}(x_i^2 - 2\mu_0 x_i + \mu_0^2)$$

Neyman-Pearson Lemma

Example con't

$$\exp -\frac{1}{2\sigma^2}(n(\mu_1^2 - \mu_0^2) - 2n\bar{x}(\mu_1 - \mu_0)) \geq k$$

Finish in class

Uniformly Most Powerful Tests

The Neyman-Pearson lemma gives us the most powerful test for a simple null hypothesis against a simple alternative hypothesis. It can be extended for composite alternative hypotheses by ensuring each simple alternative is accounted for.

Uniformly Most Powerful Test (UMP)

For a continuous random variable, let R be the critical region of size α. A test is the uniformly most powerful if it is a most powerful test against each simple alternative in the alternative (composite) hypothesis.
The critical region is called the most powerful critical region of size α.

Monotone Likelihood Ratio

We defined the Uniformly Most Powerful test, but we must state when it exists.

For $X_1, ..., X_n$ with likelihood function $L(\theta|X) = \prod_{i=1}^{n} f(X_i|\theta)$ we define:

Monotone Likelihood Ratio Property

The family (or set) of distributions has Monotone Likelihood Ratio (MLR) if we can represent the likelihood ratio as

$$\frac{f(X|\theta_2)}{f(X|\theta_1)} = f(T(X), \theta_1, \theta_2) \text{ for } \theta_2 > \theta_1$$

Where the function $f(T(X), \theta_1, \theta_2)$ is non-decreasing in $T(X)$ (strictly increasing in T(X)).

Distributions having the MLR property include exponential, binomial, normal (unknown mean, known variance), and Poisson. Any regular exponential family with $g(t(x)|\theta) = h(t)c(\theta)\exp(w(\theta)t(x))$ has a MLR if $w(\theta)$ is a non-decreasing function.

Monotone Likelihood Ratio

MLR Example

Consider $X_1, .., X_n \sim N(\mu, 1)$. The pdf is

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2}}$$

and the likelihood is

$$f(X|\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2}$$

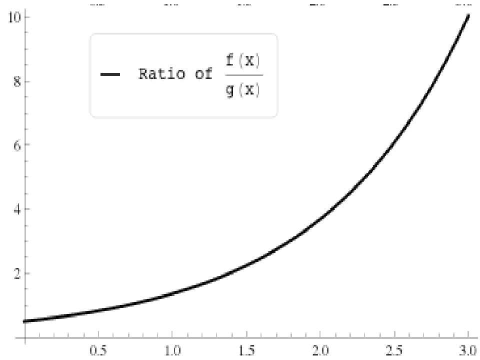
Then the likelihood ratio can be written as

$$\frac{f(X|\mu_2)}{f(X|\mu_1)} = e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_2)^2} + e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_1)^2} = e^{(\mu_2-\mu_1)\sum_{i=1}^{n}X_i-\frac{n}{2}(\mu_2^2-\mu_1^2)}$$

For $\mu_2 > \mu_1$ the likelihood ratio is increasing in $T(X) = \sum_{i=1}^{n} X_i$ and the MLR property holds.

Monotone Likelihood Ratio

The MLR property tells us that the likelihood ratio $f(X|\theta_2)/f(X|\theta_1)$ is a non-decreasing function of T(X) defined when $\theta_2 > \theta_1$



Parameter T(X) versus the ratio $f(X|\theta_2)/f(X|\theta_1)$, showing MLR property.
Source: Wikipedia

Monotone Likelihood Ratio

We link the MLR property with UMP tests:
If statistic T(X) has the MLR property then the UMP test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ or $H_1 : \theta < \theta_0$ exists and can be expressed in terms of T(X) rather than in terms of the full likelihood ratio.

Monotone Likelihood Ratio

It holds true that if T(X) has MLR property then:

1. $T(X) > t_0$ is a UMP for $H_0 : \theta \leq \theta_0$ (or $\theta = \theta_0$) vs. $H_0 : \theta > \theta_0$
2. $T(X) < t_0$ is a UMP for $H_0 : \theta \geq \theta_0$ (or $\theta = \theta_0$) vs. $H_0 : \theta < \theta_0$

Monotone Likelihood Ratio

For 1), Consider $H_0 : \theta = \theta_0$ vs $H_0 : \theta > \theta_0$.
$T(X) > t_0$ is a UMP for $H_0 : \theta \leq \theta_0$ (or $\theta = \theta_0$) vs. $H_0 : \theta > \theta_0$

Karlin-Rubin Theorem

We link sufficient statistics and the MLR property with UMP tests through the Karlin-Rubin Theorem.
Suppose we are testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. Let T(X) be a sufficient statistic, and the family of distributions for θ has MLR property in T(X). Then for any k, the test with rejection region $T(X) > k$ is the UMP test (given α) where $P(T(X) > k) = \alpha$.
By the same argument, we can apply this test to $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$ given $T(X) < k$ for the UMP α-level test where $P(T(X) < k) = \alpha$.

Non-Existence of UMP Tests

Two sided hypothesis do not have UMP tests. The reason is that the test that is UMP for $\theta < \theta_0$ is not the same as the test that is UMP for $\theta > \theta_0$. For the two sided case to exist, it would have to be most powerful across every value in the alternative hypothesis.

Interval Estimates

- We have examined point estimators of unknown distribution parameters $\theta_1, \theta_2, ..., \theta_n$ using MLE, numerical methods for MLE and MOM
- We have assessed our estimates by examining sufficiency, their bias, MSE, and MVUE.
- Even if we have minimized the squared error or have an minimum variance unbiased estimate, we have no idea if our parameter lies in an acceptable range (and where the parameter lies in that range).
- Interval estimates are calculated with our sample measurements and are two numbers that define endpoints.
- Ideally the interval has two properties: that it contains the target parameter θ and that it is relatively narrow. We want the interval to have high probability of containing θ.

Interval Estimates

Definition: Interval Estimator

Given a random sample $X_1, X_2, ..., X_n$, an interval estimate of an unknown parameter θ from probability distribution function $f(x|\theta)$ is any pair of functions $L(X_1, X_2, ..., X_n)$ and $U(X_1, X_2, ..., X_n)$ that satisfy $L(X) \leq U(X)$. $L(x)$ and $U(x)$ are the lower and upper limits of the interval, respectively.

When $X = x_1, x_2, ..., x_n$ is observed, the inference $L(x) \leq \theta \leq U(x)$ is made. $[L(x), U(x)]$ is the interval estimator.

Coverage Probability

Definition: Coverage probability

For an interval estimator $[L(x), U(x)]$, the probability that the interval contains the true parameter θ is defined by $P(\theta \in [L(x), U(x)]|\theta)$

Concept

Given a random sample $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$, we estimate μ, the expected value, with $\bar{X}$

However, there will be some estimation error between $\bar{X}$ and μ, as the probability of our estimate being exactly correct $P(\bar{X} = \mu)$ is 0. So it is more appropriate to define a range of values around $\bar{X}$ that has high probability of containing μ

We say that the probability of μ is covered by the interval $\bar{X} \pm c$ via:

$$P(\bar{X} - c \leq \mu \leq \bar{X} + c)$$

Coverage Probability

Example (CB 9.1.3)

Given a random sample $X_1, X_2, X_3, X_4 \sim N(\mu, 1)$, we estimate μ, the expected value, with $\bar{X}$
We have confidence interval $[\bar{X} - 1, \bar{X} + 1]$, and the probability that μ is covered by this interval is:

$$\begin{aligned} P(\mu \in [\bar{X} - 1, \bar{X} + 1]) &= P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) \\ &= P(-1 \leq \bar{X} - \mu \leq 1) \\ &= P\left(-2 \leq \frac{\bar{X} - \mu}{\sqrt{1/4}} \leq 2\right) \\ &= P(-2 \leq Z \leq 2) \\ &= 0.9544 \end{aligned}$$

Here, Z is the standard normal. And given these endpoints, we have over a 95% chance of covering the unknown μ with our interval estimator.

Coverage Probability and Confidence Coefficients

For an interval estimator of a parameter θ, $[L(x), U(x)]$, the confidence coefficient of $[L(x), U(x)]$ is the infimum of the coverage probabilities, $inf_\theta P_\theta(\theta \in [L(x), U(x)])$

Example (CB 9.1.6)

Example done in class

Confidence Intervals

Methods for finding interval estimators

- Interval estimators, in combination with a measure of confidence (via a confidence coefficient) are referred to as confidence intervals.
- $P(\theta \in [L(x), U(x)]|\theta) = 1 - \alpha$, where $1 - \alpha$ is the confidence coefficient
- The two primary methods for finding confidence intervals are by inverting a test statistic and the pivotal method.

Confidence Intervals: Inverting a Test Statistic

There is a strong relationship between hypothesis testing and confidence intervals in that a confidence interval can be obtained by inverting a hypothesis test (and vice versa). Specifically, the $1 - \alpha$ confidence interval is obtained by inverting the acceptance region of the α-level test.
In general, under $H_0 : \theta = \theta_0$,

$$A(\theta_0) = \{\mathbf{x} : \text{the test accepts } H_0 : \theta = \theta_0\}$$

Where $A(\theta_0)$ is the acceptance region. Defining the 1-α confidence set $C(X_1, .., X_n)$:

$$C(X_1, ..., X_n) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\}$$

The confidence interval is the set of all parameters for which the hypothesis would have accepted H_0. Namely, it is the set of θ given $X_1, .., X_n$ and for each $\theta_0 \in C(X)$ you would not reject $H_0 : \theta = \theta_0$

Confidence Intervals: Inverting a Test Statistic

Conversely we can say:

$$A(\theta_0) = \{\mathbf{x} : \theta_0 \in C(X_1, ..., X_n)\}$$

In this case $A(\theta_0)$ is the acceptance region of the α-level test of $H_0 : \theta = \theta_0$
Proof:
For the confidence set, since $A(\theta_0)$ is the acceptance region of the α-level test

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) \leq \alpha \leftrightarrow P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1 - \alpha$$

Using θ more generally than θ_0, the coverage probability of the set $C((X)$ is

$$P_{\theta_0}(\theta \in C(\mathbf{X})) = P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1 - \alpha$$

Showing that $C(\mathbf{X})$ is a $1 - \alpha$ confidence set. For the hypothesis test, the probability of Type I error for testing the null hypothesis $H_0 : \theta = \theta_0$ with acceptance region $A(\theta_0)$ is

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) = P_{\theta_0}(\theta_0 \notin C(\mathbf{X})) \leq \alpha$$

So this is the α-level test.

Confidence Intervals: Inverting a Test Statistic

In hypothesis testing, the acceptance region is the set of $X_1, .., X_n$ that are very likely for θ_0. We fix the parameter and find what sample values in the region are consistent with that value.
In interval estimation, the confidence interval is a set of θ's that make $X_1, .., X_n$ very likely.

Example: Inverting a Normal Test

Given $X_1, ..., X_n$ from a normal distribution where σ is known and we wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ for a fixed α level, we used the test statistic

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

This has critical region $\{x : |\bar{x} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}\}$, or in other words, H_0 is accepted in the region defined by $|\bar{x} - \mu_0| \leq z_{\alpha/2}\sigma/\sqrt{n}$

Confidence Intervals: Inverting a Test Statistic

Example: Inverting a Normal Test, con't

This is written as

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Which has size α, so we can write $P(H_0 \text{is rejected}|\mu = \mu_0) = \alpha$ or equivalently stated another way, $P(H_0 \text{is accepted}|\mu = \mu_0) = 1 - \alpha$. Combining this,

$$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}|\mu = \mu_0) = 1 - \alpha$$

But the probability statement is true for every μ_0, so the above is written

$$P_\mu(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

And the confidence interval $[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$ is obtained by inverting the acceptance region of the α-level test giving a $1 - \alpha$ confidence interval.

Confidence Intervals: Inverting a Test Statistic

For one sided tests, we have the following:

- L(X) is a lower confidence bound when $U(X)=\infty$
- U(X) is an upper confidence bound when $L(X)=-\infty$

Confidence Intervals: Inverting a LR Statistic

We can consider using $\lambda(X) = L(\theta_0)/L(\hat{\theta})$ as our statistic for inverting an α-level test of $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$.
We first define the LRT statistic by $\lambda(x)$ and then for a fixed θ_0 we define the acceptance region:

$$A(\theta_0) = \{x : \lambda(x) \geq k\}$$

Where k is a constant chosen to satisfy $P(X \in A(\theta_0)) = 1 - \alpha$. Since we are looking at the acceptance region, we are not splitting the region (as in UMP two sided which is non-existent). See CB Figure 9.2.2.
The associated 1-α confidence set is:

$$C(X) = \{\theta : \lambda(x) \geq k\}$$

The confidence interval can be expressed in the form:

$$\{\theta : L(\lambda(x)) \leq \theta \leq U(\lambda(x))\}$$

Where L and U are functions determined by the constraints that the set of the acceptance region has probability 1-α.

Confidence Intervals: Pivotal Method

Definition: Pivotal quantities

A random variable $Z(X,\theta) = Z(X_1, ..., X_n, \theta)$ is a pivotal quantity if the distribution of $Z(X,\theta)$ is independent of all the parameters θ
The function $Z(X,\theta)$ will usually contain both parameters and statistics, but for any set $\mathcal{A}$, $P_\theta(Z(X,\theta) \in \mathcal{A})$ cannot depend on θ.

- It is desirable to have the length U-L or mean length E(U-L) to be as short as possible.
- The pivotal method assures the minimization of the MSE of the constructed confidence interval (in most cases).

Confidence Intervals: Pivotal Method

- The pivotal method involves the following steps:
  1. Determine the point estimator for the unknown parameter θ. i.e. $\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$
  2. Construct a function of $\hat{\theta}$ and θ, $Q = g(\hat{\theta}, \theta)$ with known distribution function $f_Q(q)$ which is independent of θ and any other unknown parameter. This is the pivotal quantity.
  3. Using the distribution function $f_Q(q)$, find two constants a and b with $a < b$ such that $P(a \leq Q \leq b) = 1 - \alpha$. Usually these constants are chosen so $P(Q < a) = P(Q > b) = \alpha/2$.
  4. Put everything together into a double inequality which constructs the confidence interval, $a \leq g(\hat{\theta}, \theta) \leq b$. In terms of the unknown parameter, the equivalent inequality is $L(X) \leq \theta \leq U(X)$.
  $$P(L(X) \leq \theta \leq U(X)) = P(a \leq Q \leq b) = 1 - \alpha$$
- Note that the interval does not depend on the unknown parameter.
- When choosing a and b such that $P(a \leq Q \leq b) = 1 - \alpha$ we want the interval length $b - a$ to be as small as possible because the shorter the interval, the more precise it is.
- Typically when the distribution of Q is symmetric, the interval is also symmetric.

Confidence Intervals: Pivotal Method

In location, scale and location scale families there are many possible pivotal quantities. For $X_1, X_2, ..., X_n$ we let $\bar{X}$ and S be the sample mean and standard deviation, respectively. The common pivotal quantities are shown below.

PDF	Type	Pivotal Quantity
$f(x-\mu)$	Location	$\bar{X}-\mu$
$\frac{1}{\sigma}f(\frac{x}{\sigma})$	Scale	$\frac{\bar{X}}{\sigma}$
$\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$	Location-Scale	$\frac{\bar{X}-\mu}{S}$

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 1

Given a random sample $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$ where σ^2 is known and we wish to construct a confidence interval for μ (Location type example).

1. The estimator for the unknown parameter μ is $\bar{X}$ and it has distribution $N(\mu, \sigma^2/n)$
2. The function $Q = (\bar{X}, \mu)$ defined by
$$Q = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
has $f_Q(q) \sim N(0, 1)$ which is independent of μ
3. Using $Q \sim N(0, 1)$ we find the constants $P(Q < a) = \alpha/2$ so $a = -q_{1-\alpha/2}$ since the distribution of Q is symmetric. Similarly, $P(Q < b) = 1 - \alpha/2$ so $b = q_{1-\alpha/2}$. Note, q_α represents the upper α percentage of the standard normal distribution. In this case, $\pm q_{1-\alpha/2}$ are obtained from the standard normal distribution so we will use $\pm z_{1-\alpha/2}$.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 1 con't

4. Putting this together into a double inequality,

$$-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}$$
$$-\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \bar{X}-\mu \leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$
$$\bar{X}-\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{X}+\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$

So, the random interval $[\bar{X}-\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}+\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}]$ is an exact confidence interval for μ with confidence coefficient $1-\alpha$.
Note: The length of the interval is constant, $l = 2\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 2

Given a random sample $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$ where μ is known and σ is unknown, and we wish to construct a confidence interval for σ^2 (Scale type example).

1. The estimator for the unknown parameter σ^2 is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X - \bar{X})^2$.
2. The function $Q = (S^2, \sigma^2)$ defined by
$$Q = \frac{(n-1)S^2}{\sigma^2}$$
has $f_Q(q) \sim \chi^2_{n-1}$ which is independent of σ^2
3. Using $Q \sim \chi^2_{n-1}$ we find the constants $P(Q < a) = \alpha/2$ and $P(Q < b) = 1 - \alpha/2$.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 2 con't

4. Putting this together into a double inequality,

$$\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}$$

$$\frac{\chi^2_{n-1,\alpha/2}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{n-1,1-\alpha/2}}{(n-1)S^2}$$

$$\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}$$

So, the random interval $[(n-1)S^2/\chi^2_{n-1,1-\alpha/2}, (n-1)S^2/\chi^2_{n-1,\alpha/2}]$ is an exact confidence interval for σ^2 with confidence coefficient $1-\alpha$.
Note, this interval is not symmetric since the χ^2 distribution is not symmetric.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 3

Given a random sample $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$ where μ and σ^2 are unknown, and we wish to construct a confidence interval for μ and σ^2 (Location-Scale type example).

1. The estimator for the unknown parameter μ is $\bar{X}$ and it has distribution $N(\mu, \sigma^2/n)$
2. As in the previous example, the function $Q = (\bar{X}, \mu)$ defined by

   $$Q = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

   has $f_Q(q) \sim N(0,1)$ which is independent of μ, but note that it also contains the unknown parameter σ. The confidence interval cannot contain an unknown parameter, so we must replace it with the unbiased estimator for σ^2, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X - \bar{X})^2$.

   The pivotal function becomes, $Q = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ which has a t_{n-1} distribution.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 3 con't

3. Using $Q \sim t_{n-1}$ we find the constants $P(Q < a) = \alpha/2$ or $P(Q < a) = -(1 - \alpha/2)$, so $a = -t_{1-\alpha/2,n-1}$ since the t-distribution is symmetric. Similarly, $P(Q < b) = 1 - \alpha/2$ so $b = t_{1-\alpha/2,n-1}$.
4. Putting this together into a double inequality,

$$
\begin{aligned}
-t_{1-\alpha/2,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{1-\alpha/2,n-1} \\
-\frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} \\
\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1}
\end{aligned}
$$

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 3 con't

So, the random interval $[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1}, \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1}]$ is an exact confidence interval for μ with confidence coefficient $1 - \alpha$.

When the sample size is large ($n \geq 35$) the CLT states that the t-distribution is approximated by the standard normal distribution. Thus in this case, $t_{1-\alpha/2,n-1} = z_{1-\alpha/2}$ and the confidence interval for μ becomes $[\bar{X} - \frac{S}{\sqrt{n}} z_{1-\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}} z_{1-\alpha/2}]$. This is an asymptotic result (Ch. 10).

Confidence Intervals: Pivotal Method

Practical Example: CI for μ given

Given a random sample $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$ of $n = 14$ gym-goers showed that the mean workout time was $\bar{X} = 45$ minutes with a sample standard deviation of $s = 14$ minutes. What is the population mean μ with confidence coefficient $1 - \alpha = 0.95$?

- The confidence coefficient $1 - \alpha = 0.95$ means $\alpha = 0.05$ and thus $\alpha/2 = 0.025$
- Use $Q \sim t_{n-1}$ as the pivot
- Using the table of standard normals, $t_{0.025,13} = 2.16$ and therefore

$$\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} = 45 - \frac{14}{\sqrt{14}} 2.16 = 36.92$$
$$\bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} = 45 + \frac{14}{\sqrt{14}} 2.16 = 53.08$$

- Therefore the CI for the average workout time (μ) with $1 - \alpha = 0.95$ (also stated as $100(1 - \alpha) = 95\%$ confidence) is $[36.92, 53.08]$ minutes.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 3 con't

1. The estimator for the unknown parameter σ^2 is S^2
2. The function $Q = (S^2, \sigma^2)$ defined by
$$Q = \frac{(n-1)S^2}{\sigma^2}$$
has $f_Q(q) \sim \chi^2_{n-1}$
3. The constants a and b are obtained with the χ^2_{n-1} distribution
$$P(\frac{(n-1)S^2}{\sigma^2} > a) = 1 - \alpha/2$$
$$a = \chi^2_{1-\alpha/2,n-1}$$
$$P(\frac{(n-1)S^2}{\sigma^2} > b) = \alpha/2$$
$$b = \chi^2_{\alpha/2,n-1}$$
4. Putting this together, we get the confidence interval for σ^2
$$\frac{(n-1)S^2}{\sqrt{\chi^2_{\alpha/2,n-1}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\sqrt{\chi^2_{1-\alpha/2,n-1}}}$$
5. The confidence intervals for μ and σ^2 are constructed separately in this case.

Confidence Intervals: Pivotal Method

Pivotal method for Normal: Case 3 con't

Suppose we wish to find a simultaneous confidence interval for μ and σ^2.
In this situation one good option is to use the Bonferroni inequality (CB 1.2.9).
Recall $P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$. The probability that the interval covers μ is $P(A_1) = 1 - \alpha/2$, and similarly for σ^2, $P(A_2) = 1 - \alpha/2$. So, the inequality gives $2(1 - \alpha/2 - 1 = 1 - \alpha$.
We thus can use the same pivots shown above for μ and σ^2, however we require a and b to be defined by $\pm t_{n-1,\alpha/4}$ for μ and $a = \chi^2_{n-1,1-\alpha/4}, b = \chi^2_{n-1,\alpha/4}$ for σ^2.
The simultaneous $1 - \alpha$ CI for (μ, σ^2) is

$$\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/4,n-1} \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/4,n-1}, \frac{(n-1)S^2}{\sqrt{\chi^2_{\alpha/4,n-1}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\sqrt{\chi^2_{1-\alpha/4,n-1}}}$$

Confidence Intervals: Pivoting the CDF

The pivots, Q, defined previously were constructed with location and scale methods. We saw that they were straightforward to find for the normal distribution, and in CB (9.2.8) we see the use of the Gamma pivot.
Another, more general pivot is using the CDF. We present this method for the continuous CDF:

Pivoting the continuous CDF

Assume T is a statistic with continuous CDF $F_T(t|\theta) = P(T \leq t)$. Given $\alpha_1 + \alpha_2 = \alpha, 0 < \alpha < 1$ (often $\alpha_1 = \alpha_2$), for each $t \in T$:

1) When $F_T(t|\theta)$ is a decreasing function of θ for each t, $\theta_L(t)$ and $\theta_U(t)$ can be defined by $F_T(t|\theta_U(t)) = \alpha_1$ and $F_T(t|\theta_L(t)) = 1 - \alpha_2$.

2) When $F_T(t|\theta)$ is an increasing function of θ for each t, $\theta_L(t)$ and $\theta_U(t)$ can be defined by $F_T(t|\theta_U(t)) = 1 - \alpha_2$ and $F_T(t|\theta_L(t)) = \alpha_1$.

Evaluating Interval Estimators

In evaluating the confidence intervals we have generated through the means described above, we look at two related quantities: length (size) and coverage probability. Interval length and coverage probability vie against each other. We discuss coverage probability previously, so here focus on defining the length of an interval.

Length: We choose the $1-\alpha$ interval such that is as short as possible through minimizing the expectation of $U(X) - L(X)$. It is important to note:

- when α is fixed, if n increases the confidence interval length decreases. So smaller n will result in larger intervals.
- when n is fixed, if $1-\alpha$ increases, a and b increase, and so will the confidence interval. Smaller $1-\alpha$ results in smaller a and b and thus a smaller interval.

Theorem: Interval Length

Let f(x) be a unimodal pdf. If the interval [a,b] satisfies

1. $\int_a^b f(x)dx = 1-\alpha$
2. $f(a) = f(b) > 0$
3. $a \leq x* \leq b$ where x* is the mode of f(x)

Then [a,b] is the shortest among all intervals that satisfy 1.

Evaluating Interval Estimators: Optimizing Length

Suppose we have $X_1, ..., X_n \sim N(\mu, \sigma^2)$ where both parameters are unknown. Find the shortest length $1-\alpha$ confidence interval for μ using a pivotal quantity.
Solution:
We saw before that our pivot for μ in this case is $Q = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ which has a t_{n-1} distribution. We let

$$P(a \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq b) = 1-\alpha$$
$$P(\bar{X} - bS/\sqrt{n} \leq \mu \leq \bar{X} - aS\sqrt{n}) = 1-\alpha$$

So the length of the interval is $L = (b-a)/S/\sqrt{n}$. But what choice of a and b is best?

Evaluating Interval Estimators: Optimizing Length

We wish to find a and b such that L is minimized subject to

$$\int_a^b f(x)dx = 1-\alpha$$

We can see that

$$\frac{dL}{da} = (\frac{db}{da} - 1)\frac{S}{\sqrt{n}}$$

$$f(b)\frac{db}{da} - f(a) = 0$$

$$\text{so } \frac{dL}{da} = (\frac{f(a)}{f(b)} - 1)\frac{S}{\sqrt{n}}$$

The minimum occurs at a=-b. We use $a = -b = -t_{n-1,\alpha/2}$ to get the shortest length $1-\alpha$ confidence interval for the pivotal quantity $\sqrt{n}(\bar{X}-\mu)/S$