Testing Hypothesis

Maura Mezzetti

Department of Economics and Finance
Università Tor Vergata

Hypothesis Testing

"*It is a mistake to confound strangeness with mystery*"

Sherlock Holmes
A Study in Scarlet

Outline

General Concept

Definition

A hypothesis is a statement about a population parameter

For example:

- population mean
- population proportion
- population variance
- distribution

General Concept

Definition

A hypothesis is a statement about a population parameter

For example:

- population mean
- population proportion
- population variance
- distribution

General Concept

Hypothesis need to be translated in a statement:

For example:

- population mean: $H_0 : \mu = 0 \quad \textit{versus}\ H_1 : \mu \neq 0$
- population proportion: $H_0 : \pi \geq 0.5 \quad \textit{versus}\ H_1 : \mu < 0.5$

General Concept

Hypothesis need to be translated in a statement:

For example:

- population mean: $H_0 : \mu = 0$ *versus* $H_1 : \mu \neq 0$
- population proportion: $H_0 : \pi \geq 0.5$ *versus* $H_1 : \mu < 0.5$

General Concept

Definition

A hypothesis testing is an inferential statistical procedure used to offer evidence of a hypothesis

A test of a hypothesis is a process that uses sample statistics to test a claim about the value of a population parameter and to decide, which of two complementary hypothesis in a hypothesis testing problem is true

General Concept

Definition

A hypothesis testing is an inferential statistical procedure used to offer evidence of a hypothesis

A test of a hypothesis is a process that uses sample statistics to test a claim about the value of a population parameter and to decide, which of two complementary hypothesis in a hypothesis testing problem is true

What's a Hypothesis?

Two complementary hypothesis in a hypothesis testing problem are called the null and the alternative hypothesis

The Null Hypothesis

indicated with H_0, is a hypothesis that is held to be true unless sufficient evidence to the contrary is obtained

What's a Hypothesis?

Two complementary hypothesis in a hypothesis testing problem are called the null and the alternative hypothesis

The Null Hypothesis

indicated with H_0, is a hypothesis that is held to be true unless sufficient evidence to the contrary is obtained

Alternative Hypothesis

indicated with H1, is a new hypothesis to be tested. A hypothesis against which the null hypothesis is tested and which will be held to be true if the null is held false.

What's a Hypothesis?

Two complementary hypothesis in a hypothesis testing problem are called the null and the alternative hypothesis

The Null Hypothesis

indicated with H_0, is a hypothesis that is held to be true unless sufficient evidence to the contrary is obtained

Alternative Hypothesis

indicated with H1, is a new hypothesis to be tested. A hypothesis against which the null hypothesis is tested and which will be held to be true if the null is held false.

What's a Hypothesis?

Simple Hypothesis

Hypothesis determines completely distribution of X, specifies only one possible distribution for the sample

Composite Hypothesis

hypothesis specifies more that one possible distributions for the sample.

What's a Hypothesis?

Simple Hypothesis

Hypothesis determines completely distribution of X, specifies only one possible distribution for the sample

Composite Hypothesis

hypothesis specifies more that one possible distributions for the sample.

What's a Hypothesis?

Simple Hypothesis

- $H_0\ \mu = \mu_0$

Composite Hypothesis

- $H_0\ \mu > \mu_0$
- $H_0\ \mu < \mu_0$
- $H_0\ \mu \neq \mu_0$

What's a Hypothesis?

Simple Hypothesis

- $H_0 \, \mu = \mu_0$

Composite Hypothesis

- $H_0 \, \mu > \mu_0$
- $H_0 \, \mu < \mu_0$
- $H_0 \, \mu \neq \mu_0$

What's a Hypothesis Test?

Definition A hypothesis procedure or hypothesis test is a rule that specifies:

- For which sample values the decision is made to accept H_0 as true
- For which sample values H_0 is rejected and H_1 is accepted as true

Rejection Region

The subset of the sample space for which H_0 will be rejected is called the rejection region or critical region. The complement of the rejection region is called the acceptance region.
The rejection region R of a hypothesis test is usually defined by a test statistic $W(X)$, a function of the sample

$$R = \{X : W(X) > c\} \quad \Rightarrow \quad \textit{reject } H_0$$

$$A = \{X : W(X) \leq c\} \quad \Rightarrow \quad \textit{accept } H_0$$

Statistical rule

Procedure

Given a random sample $X_1, X_2, \ldots, X_n$ a test statistic W is a function of the random sample $W(X) = W(X_1, X_2, \ldots, X_n)$ such that

- if $W \in R$ then H_0 is rejected
- if $W \in A$ then H_0 is not rejected

R is called critical or rejection region.

Types of Errors

- Since the decision is based on a sample and not on the entire population, there is always the possibility you'll make the wrong decision.
- There are two types of errors.
  - A type I error occurs if H_0 is rejected when it is actually true.
  - A type II error occurs if H_0 is not rejected when it is actually false.

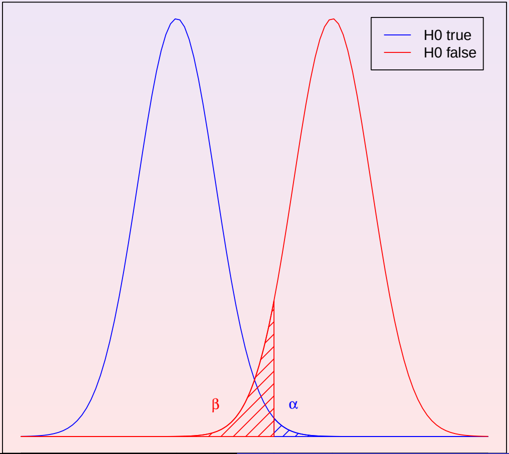
Types of Errors

Type I and II Errors

Decision→ Reality↓	H_0 is rejected	H_0 is not rejected
H_0 true	type I error α	no error $1-\alpha$
H_0 false	no error $1-\beta$	type II error β

Type I and II errors

Let H_0: $\theta = \theta_0$ vs H_1: $\theta = \theta_1$. Reject H_0 if $W > c$



Errors in making Decisions

Type I Error

- Reject true null hypothesis
- $\alpha = P(\text{reject } H_0 | H_0 \text{ true})$
- Has serious consequences: type I error is considered more dangerous, we do our best to not commit type I error
- Probability of Type I Error is denoted with α
- Called **Level of Significance**

Errors in making Decisions

Type II Error

- Do not reject false null hypothesis
- $\beta = P(\text{not reject } H_0 | H_0 \text{ false})$
- It is considered less serious
- Probability of Type II Error is denoted with β

Remarks

- The significance level of the test is fixed. Usually $\alpha = 0.01$ or $\alpha = 0.05$. Sometimes also $\alpha = 0.10$ is accepted
- Once α is fixed, the critical region C is determined, and C depend on α
- Because only the type I error is controlled by a test of significance, the size of the type II error may be large
- For that reason a test is useful only only if it reject H_0. If the test accept the null hypothesis it is useless because the type II error may be large

Tests and Trials

Trials

In the courtroom, juries must make a decision about the guilt or innocence of a defendant. Suppose you are on the jury in a murder trial. It is obviously a mistake if the jury claims the suspect is guilty when in fact he or she is innocent. What is the other type of mistake the jury could make? Which is more serious?

Decision Results

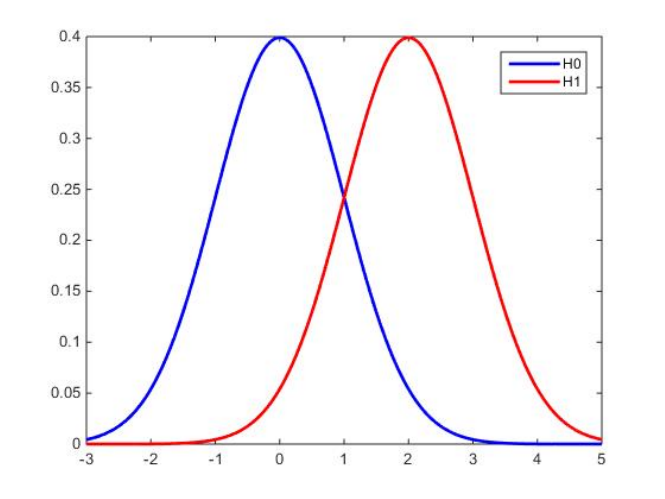
Jury Trial			H_0 Test		
Verdict	Actual Situation		Decision	Actual Situation	
	Innocent	Guilty		H_0 True	H_0 False
Innocent	Correct	Error	Accept H_0	$1 - \alpha$	Type II Error (β)
Guilty	Error	Correct	Reject H_0	Type I Error (α)	Power ($1 - \beta$)

Example $N(\mu, \sigma = 10)$, n=100

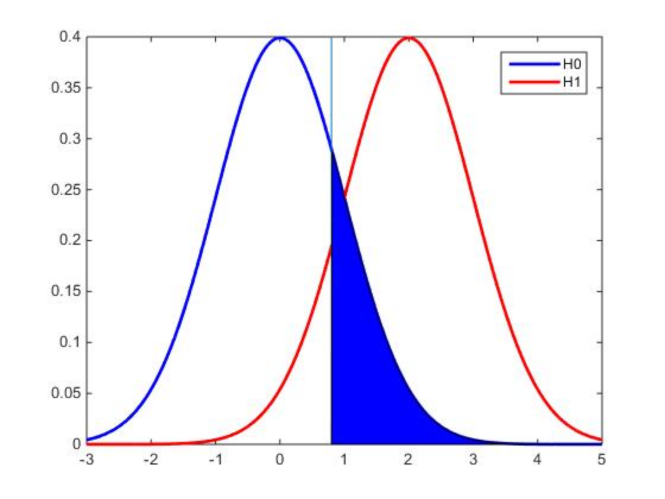
$$H_0 : \mu = 0 \qquad vs \qquad H_1 : \mu = 2$$

$$R = \{x : \bar{x} > 0.8\}$$

Example $N(\mu, \sigma = 10)$, n=100



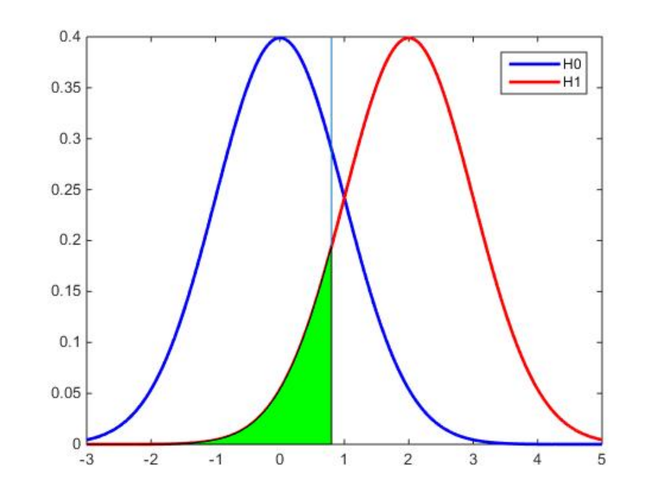
Example $N(\mu, \sigma = 10)$, n=100, α



Example $N(\mu, \sigma = 10)$, n=100, α

$$
\begin{aligned}
\alpha &= P(\bar{X} > 0.8 | H_0) \\
\alpha &= P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \frac{0.8 - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) \\
\alpha &= P\left(\frac{\bar{X} - 0}{1} > 0.8\right) \\
\alpha &= 0.2119
\end{aligned}
$$

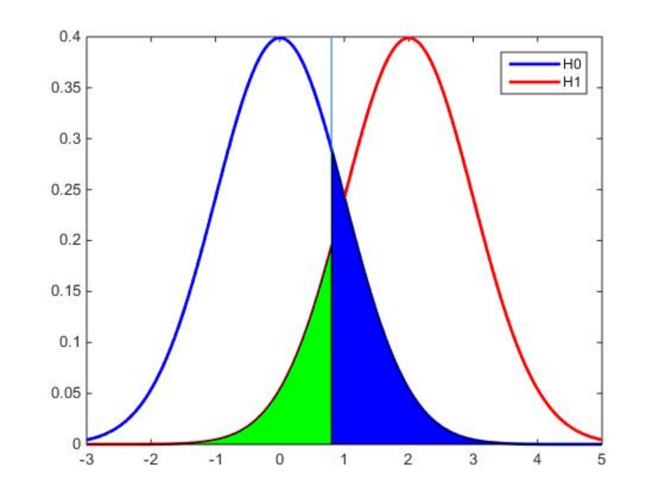
Example $N(\mu, \sigma = 10)$, n=100, β



Example $N(\mu, \sigma = 10)$, n=100, β

$$\begin{aligned}
\beta &= P(\bar{X} \leq 0.8 | H_1) \\
\beta &= P\left(\frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} \leq \frac{0.8 - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) \\
\beta &= P\left(\frac{\bar{X} - 2}{1} \leq -1.2\right) \\
\beta &= 0.1151
\end{aligned}$$

Example $N(\mu, \sigma = 10)$



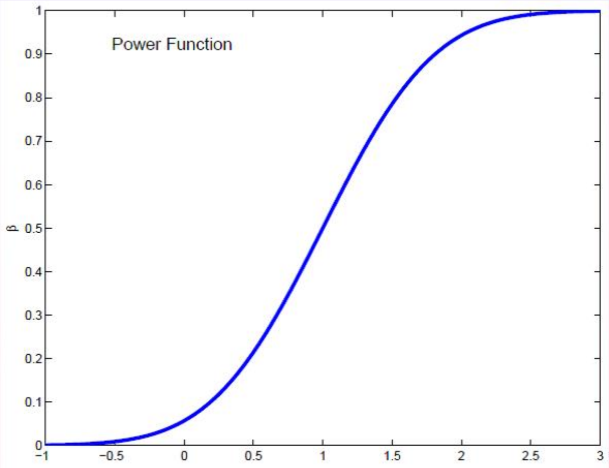
The Power Function

Definition: The *power function* of a hypothesis test with rejection region R is the function of θ defined by: $\beta(\theta) = P_\theta(\mathbf{X} \in \mathcal{R})$

Example: Let $X_1, \ldots, X_{10}$ be a random sample from a population $N(\mu, \sigma^2 = 4)$ and let $H_0 : \mu \leq 0$ and $H_1 : \mu > 0$ and let consider a test that reject H_0 when $\bar{X} > 1$, the power function has the following expression:

$$\beta(\mu) = P(\bar{X} > 1 | \mu = \mu) = P\left(Z > \frac{1-\mu}{\frac{2}{\sqrt{10}}}\right)$$

The Power Function



α and β

Definition For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test if $sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Definition For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if $sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$

Example: Computation of α and β

Let $X_1, \ldots, X_3$ be a random sample from a population $N(\mu, \sigma^2 = 4)$ and let $H_0 : \mu = 0$ and $H_1 : \mu = 1$ and let consider a test with the following rejection region:

$$R = \{(x_1, x_2, x_3) : x_1 + x_2 - x_3 > 0.8\}$$

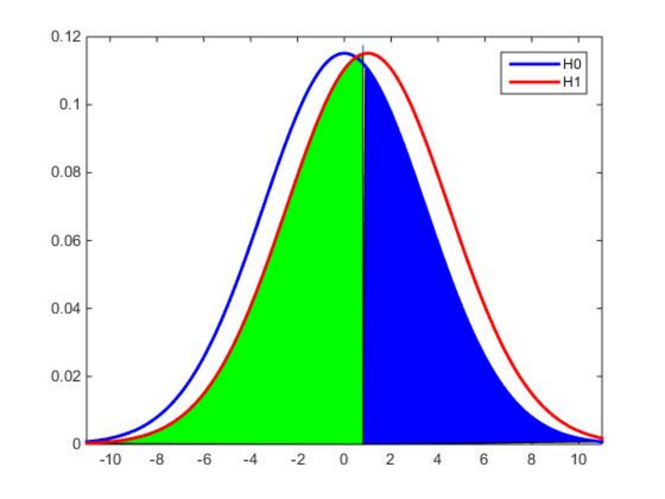
Calculate α and β

α and β

$\alpha = P(X_1 + X_2 - X_3 > 0.8|\mu = 0)$ and
$\beta = P(X_1 + X_2 - X_3 \leq 0.8|\mu = 1)$
$X_1 + X_2 - X_3 \sim N(\mu, 3 \times \sigma^2)$
Let $Y = X_1 + X_2 - X_3$

- Under null hypothesis: $Y \sim N(0, 12)$
- Under alternative hypothesis: $Y \sim N(1, 12)$

Example



α and β

$X_1 + X_2 - X_3 \sim N(\mu, 3 \times \sigma^2)$
Let $Y = X_1 + X_2 - X_3$

- Under null hypothesis: $Y \sim N(0, 12)$
- Under alternative hypothesis: $Y \sim N(1, 12)$

$\alpha = P(Y > 0.8|\mu = 0) = P\left(\frac{Y-0}{\sqrt{12}} > \frac{0.8-0}{\sqrt{12}}\right) = P\left(Z > 0.23\right) = 0.41$
$\beta = P(Y \leq 0.8|\mu = 1) = P\left(\frac{Y-1}{\sqrt{12}} \leq \frac{0.8-1}{\sqrt{12}}\right) = P\left(Z \leq -0.058\right) = 0.477$

Example: Computation of α and β

Let $X_1, \ldots, X_3$ be a random sample from a population $N(\mu, \sigma^2 = 4)$ and let $H_0 : \mu = 0$ and $H_1 : \mu = 1$ and let consider a test with the following rejection region:

$$R = \left\{ (x_1, x_2, x_3) : \frac{x_1 + x_2 + x_3}{3} > 0.8 \right\}$$

Calculate α and β

α and β

$\alpha = P\left(\frac{X_1+X_2+X_3}{3} > 0.8|\mu = 0\right)$ and

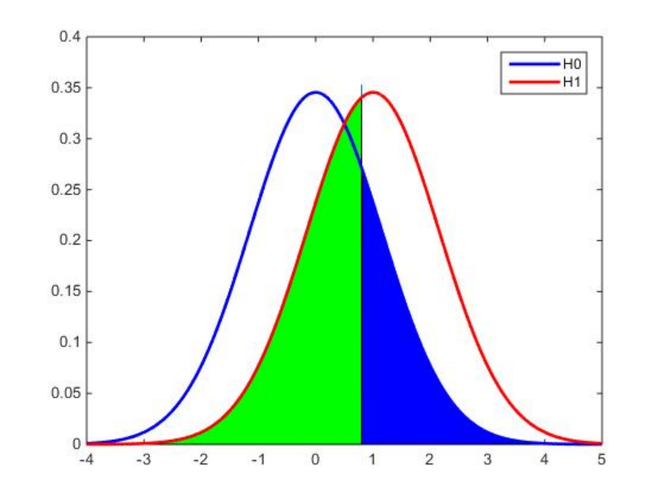
$\beta = P\left(\frac{X_1+X_2+X_3}{3} \leq 0.8|\mu = 1\right)$

$\frac{X_1+X_2+X_3}{3} \sim N(\mu, \frac{\sigma^2}{3})$

Let $\frac{X_1+X_2+X_3}{3}$

- Under null hypothesis: $Y \sim N\left(0, \frac{4}{3}\right)$
- Under alternative hypothesis: $Y \sim N\left(1, \frac{4}{3}\right)$

Example



α and β

$$\alpha = P(Y > 0.8|\mu = 0) = P\left(\frac{Y-0}{\sqrt{4/3}} > \frac{0.8-0}{\sqrt{4/3}}\right) = 0.2442$$

$$\beta = P(Y \leq 0.8|\mu = 1) = P\left(\frac{Y-1}{\sqrt{4/3}} \leq \frac{0.8-1}{\sqrt{4/3}}\right) = 0.4312$$

Most Powerful Tests

Definition: The region C is a uniformly most powerful critical region of size α for testing the simple hypothesis H_0 against a composite alternative hypothesis H_1 if C is a best critical region of size α for testing H_0 against each simple hypothesis in H_1. The resulting test is said to be uniformly most powerful.

Most Powerful Tests

Definition: Let C be a class of tests for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ of size α. A test in class C, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class C test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and for every $\beta'(\theta)$ that is a power function of a test in class C.

Evaluating a Test

Size of a Test

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size-α test if $sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Uniformly Most Powerful Test of size α

Let C be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ of size α. A test in class C, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class C test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_1$ and for every $\beta'(\theta)$ that is a power function of a test in class C.

Evaluating a Test

Size of a Test

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size-α test if $sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Uniformly Most Powerful Test of size α

Let C be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ of size α. A test in class C, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class C test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_1$ and for every $\beta'(\theta)$ that is a power function of a test in class C.

Neyman-Pearson Lemma

Theorem: Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ where the pdf or pmf corresponding to θ_i is $f(x|\theta_i)$, $i = 0, 1$ using a test with rejection region R that satisfies

$$x \in R \quad \textit{if } f(x|\theta_1) > kf(x|\theta_0)$$

and

$$x \in R^C \quad \textit{if } f(x|\theta_1) \leq kf(x|\theta_0)$$

and

$$\alpha = P_{\theta_0}(X \in R)$$

Any test that satisfies previous conditions is a UMP level α test.

Corollary of Neyman-Pearson

Corollary: Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Suppose $T(\mathbf{X})$ is a sufficient statistics for θ and $g(x|\theta_i)$, $i = 0, 1$ is the pdf or pmf corresponding to θ_i. Then any test statistics based on T with rejection region S is a UMP level α test if it satisfies

$$t \in S \quad \textit{if } g(t|\theta_1) > kg(t|\theta_0)$$

and

$$t \in S^C \quad \textit{if } g(t|\theta_1) \leq kg(t|\theta_0)$$

and

$$\alpha = P_{\theta_0}(T \in S)$$

Application Neyman-Pearson Lemma

- Suppose X has an exponential distribution with arrival rate λ. We wish to test the hypothesis that $\lambda = 1$ against the alternative that $\lambda = 2$.
- Suppose X has a Gaussian distribution with mean μ and known variance. We wish to test the hypothesis that $\mu = \mu_0$ against the alternative that $\mu = \mu_1$.

LRT and UMP test

Theorem (UMP tests) If the LR test of size α for $H_0 : \theta = \theta_0$ versus the simple alternative $H_0 : \theta = \theta_1$ is the same test for all $\theta_1 \in \Theta_1$, (for example $\Theta_1 = (\theta_1, \infty)$) then this is the UMP test of size α for $H_0 : \theta = \theta_0$ against the compound alternative $H_0 : \theta \in \Theta_1$. Otherwise no UMP test of these hypotheses of size α exists.

Remark So whenever an UMP test exists for a simple H_0 it is a LR test.

Generalized Likelihood Ratio Tests

We have been looking at really simple problems, typically with just one parameter, but in the real world of applied statistics we constantly deal with models having many parameters. In that context we never have a simple null hypothesis. Our null hypotheses often assert that one of the parameters takes a specific value, but such a hypothesis is not simple, since that would entail specifying the values of all the parameters.

Likelihood Ratio Test

Definition: The likelihood ratio test statistics for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^C$ is

$$\lambda(\mathbf{x}) = \frac{sup_{\Theta_0} L(\theta|\mathbf{x})}{sup_{\Theta} L(\theta|\mathbf{x})}, \qquad \Theta = \Theta_0 \cup \Theta_0^C$$

Definition: The likelihood ratio test (LRT) is any test that has a rejection region of the form : $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ where c is any number satisfying $0 \leq c \leq 1$.

This should be reduced to the simplest possible form.

Likelihood Ratio Tests

The rationale behind LRTs may best be understood in the situation in which $f(x|\theta)$ is the probability distribution of a discrete random variable. In this case, the numerator of $\lambda(\mathbf{x})$ is the maximum probability of the observed sample, maximum being computed under the null hypothesis. On the other hand, the denominator of $\lambda(\mathbf{x})$ is the maximum probability of the observed sample over all possible parameters.

Likelihood Ratio Tests

The ratio of these two maxima is small if there are parameter points in H_1 for which the observed sample is much more likely than for any parameter in H_0. In this situation, the LRT criterion says H_0 should be rejected and H_1 accepted as true.

Relation between LRT and MLE

Suppose MLE of θ exists. Let $\hat{\theta}_0$ be the MLE of θ in the null set Θ_0 (restricted maximization) and $\hat{\theta}$ be the MLE of θ in the full set Θ (unrestricted maximization). Then the LRT statistic, a function of x (not θ) is

$$\lambda(\mathbf{x}) = \frac{sup_{\Theta_0} L(\theta|\mathbf{x})}{sup_{\Theta} L(\theta|\mathbf{x})} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

In $R\{x : \lambda(x) \leq c\}$, different c gives different rejection region and hence different tests.

LRT and sufficiency

Theorem If $T(X)$ is a sufficient statistic for θ, $\lambda^\star(t)$ is the LRT statistic based on T, and $\lambda(x)$ is the LRT statistic based on x. Then

$$\lambda^\star(T(x)) = \lambda(x)$$

for every x in the sample space.

Thus the simplified expression for $\lambda(x)$ should depend on x only through $T(x)$ if $T(X)$ is a sufficient statistic for θ.

Example: Likelihood ratio test

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential population with pdf:

$$f(x|\theta) = \begin{cases} exp\left(-(x-\theta)\right) & x \geq \theta \\ 0 & x < \theta \end{cases}$$

The likelihood function is:

$$L(\theta|x) = \begin{cases} exp\left(-\sum_i x_i + n\theta\right) & \theta \leq x_{(1)} \\ 0 & \theta > x_{(1)} \end{cases}$$

Example: Likelihood ratio test

Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ where θ_0 is a value specified. The likelihood function is an increasing function of θ on $-\infty < \theta \leq x_{(1)}$. The likelihood ratio test statistics is

$$\lambda(\mathbf{x}) = \begin{cases} 1 & x_{(1)} \leq \theta_0 \\ exp\left(-n(x_{(1)} - \theta_0)\right) & x_{(1)} > \theta_0 \end{cases}$$

An LRT, a test that rejects H_0 if $\lambda(\mathbf{X}) \leq c$, is a test with rejection region $\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{-logc}{n}\}$. Note that the rejection region depends on the sample only through the sufficient statistics $X_{(1)}$.

Likelihood Ratio Tests and Newman Pearson lemma

The Neyman-Pearson Lemma asserts that, in general a best critical region can be found by finding the $n-$dimensional points in the sample space for which the likelihood ratio is smaller than some constant.

Monotone Likelihood Ratio (MLR)

Definition: A family of pdfs or pmfs $\{g(t;\theta);\ \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a monotone likelihood ratio (MLR) if, for every $\theta_2 > \theta_1$, $g(t,\theta_2)/g(t,\theta_1)$ is an increasing function of t on $\{t : g(t,\theta_1) > 0 \text{ or } g(t,\theta_2) > 0\}$.

Monotone Likelihood Ratio (MLR)

Theorem Suppose that X has density $f(x;\theta)$ with MLR in $T(x)$, then:

1. There exists a UMP level α test of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ which is of the form $C(x) = \{T(x) \geq k\}$
2. $\beta(\theta)$ is increasing in θ
3. For all θ' this same test is the UMP level $\alpha' = \beta(\theta')$ to test $H_0 : \theta \leq \theta'$ versus $H_1 : \theta > \theta'$
4. For all $\theta < \theta_0$, the test of (1) minimizes $\beta(\theta)$ among all tests of size α

$N(\mu, \sigma^2)$, with σ^2 known

$H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 \qquad \mu_0 > \mu_1$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -z_{1-\alpha} \right\}$$

$H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -z_{1-\alpha} \right\}$$

$H_0 : \mu > \mu_0$ vs $H_1 : \mu < \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -z_{1-\alpha} \right\}$$

$N(\mu, \sigma^2)$, with σ^2 known

$H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 \qquad \mu_0 < \mu_1$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_{1-\alpha} \right\}$$

$H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_{1-\alpha} \right\}$$

$H_0 : \mu < \mu_0$ vs $H_1 : \mu > \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq z_{1-\alpha} \right\}$$

$N(\mu, \sigma^2)$, with σ^2 known

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_1$

$$R = \left\{ \left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq z_{1-\alpha/2} \right\}$$

$N(\mu, \sigma^2)$, with σ^2 unknown

$H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 \qquad \mu_0 > \mu_1$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}^{n-1} \right\}$$

$H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}^{n-1} \right\}$$

$H_0 : \mu > \mu_0$ vs $H_1 : \mu < \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}^{n-1} \right\}$$

$N(\mu, \sigma^2)$, with σ^2 unknown

$H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 \qquad \mu_0 < \mu_1$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha}^{n-1} \right\}$$

$H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha}^{n-1} \right\}$$

$H_0 : \mu < \mu_0$ vs $H_1 : \mu > \mu_0$

$$R = \left\{ \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha}^{n-1} \right\}$$

$N(\mu, \sigma^2)$, with σ^2 unknown

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_1$

$$R = \left\{ \left| \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \right| \geq t^{n-1}_{1-\alpha/2} \right\}$$

Method of inversion

One to one correspondence between tests and confidence intervals.

Hypothesis testing:

Fix the parameter: asks what sample values (in the appropriate region) are consistent with that fixed value

Confidence set:

Fix the sample value: asks what parameter values make this sample most plausible

Method of inversion

One to one correspondence between tests and confidence intervals.

Hypothesis testing:

Fix the parameter: asks what sample values (in the appropriate region) are consistent with that fixed value

Confidence set:

Fix the sample value: asks what parameter values make this sample most plausible

For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test $H_0 : \theta = \theta_0$. Define a set $C(x) = \{\theta_0 : x \in A(\theta_0)\}$. Then $C(x)$ is a $(1-\alpha)$-confidence set.

Method of inversion

One to one correspondence between tests and confidence intervals.

Hypothesis testing:

Fix the parameter: asks what sample values (in the appropriate region) are consistent with that fixed value

Confidence set:

Fix the sample value: asks what parameter values make this sample most plausible

For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test $H_0 : \theta = \theta_0$. Define a set $C(x) = \{\theta_0 : x \in A(\theta_0)\}$. Then $C(x)$ is a $(1 - \alpha)$-confidence set.

Hypothesis Testing and Confidence Intervals

In general, inverting acceptance region of a two sided test will give two sided interval and inverting acceptance region of a one sided test will give an open end interval on one side.

Theorem

Let acceptance region of a two sided test be of the form $A(\theta) = \{x : c_1(\theta) \leq T(x) \leq c_1(\theta)\}$ and let the cutoff be symmetric, that is, $P_\theta(T(X) > c_2(\theta)) = \alpha/2$ and $P_\theta(T(X) < c_1(\theta)) = \alpha/2$. If T has MLR property then both $c_1(\theta)$ and $c_2(\theta)$ are increasing in θ.