Econ 6190: Econometrics I

Hypothesis Testing

Chen Qiu

Cornell Economics

2024 Fall

Contents

Reference

- Hansen Ch. 13

1. Basic Concepts

Set-up

- A random vector X has distribution $F(x)$
- We are interested in (scalar) parameter θ determined by $F \in \mathscr{F}$
- The parameter space is $\theta \in \Theta$
- We have a random sample $\{X_1, X_2 \ldots X_n\}$ from distribution F
- In previous sections, we talked abut estimation of θ
- In this section, we are interested in testing some hypothesis about θ

Hypotheses

- A **hypothesis** is a statement about population parameter θ
- We call the hypothesis to be tested the null hypothesis
- **Definition**: The **null hypothesis** $\mathbb{H}_0$, is the restriction $\theta = \theta_0$ for some specific value θ_0, or $\theta \in \Theta_0$ for some subset Θ_0 of Θ. The null hypothesis is often written as

$$\mathbb{H}_0 = \{\theta \in \Theta : \theta = \theta_0\} \text{ or } \mathbb{H}_0 = \{\theta \in \Theta : \theta \in \Theta_0\}$$

- The complement of null hypothesis is alternative hypothesis
- **Definition**: The **alternative hypothesis** is the set

$$\mathbb{H}_1 = \{\theta \in \Theta : \theta \neq \theta_0\} \text{ or } \mathbb{H}_0 = \{\theta \in \Theta : \theta \notin \Theta_0\}$$

Point hypotheses

- In this note, we focus on **point hypothesis**

$$\mathbb{H}_0 = \{\theta \in \Theta : \theta = \theta_0\}$$

- The alternative hypothesis could be
  - one sided: $\mathbb{H}_1 : \theta > \theta_0$ or $\mathbb{H}_1 : \theta < \theta_0$
  - two sided: $\mathbb{H}_1 : \theta \neq \theta_0$
- One sided alternative arises if the null lies on the boundary of the parameter space $\Theta = \{\theta : \theta \geq \theta_0\}$
  - Example: some policy with non-negative effect

- A hypothesis is a restriction on the underlying distribution F
- Define the **null distribution** as a set F_0 such that

$$F_0 = \{F \in \mathscr{F} : \mathbb{H}_0 \text{ is true}\}$$

- F_0 can be a singleton (a single distribution), a parametric family, or a nonparametric family
- Suppose $\mathbb{H}_0 = \{\mu = \mu_0\}$. Examples of F_0
    - singleton: $X \sim \mathrm{N}(\mu, \sigma^2)$ with known σ^2
    - parametric: $X \sim \mathrm{N}(\mu, \sigma^2)$ with unknown σ^2
    - nonparametric: X has finite mean

Simple vs. composite hypothesis

- **Definition**: A hypothesis $\mathbb{H}$ (could be null or alternative) is **simple** if the set $\{F \in \mathscr{F} : \mathbb{H} \text{ is true}\}$ is a singleton.

  A hypothesis $\mathbb{H}$ is **composite** if the set $\{F \in \mathscr{F} : \mathbb{H} \text{ is true}\}$ contains multiple distributions

- Suppose $\mathbb{H}_0 = \{\mu = \mu_0\}$. Examples of F_0
  - singleton: $X \sim \mathrm{N}(\mu, \sigma^2)$ with known σ^2
    $\Rightarrow$ simple
  - parametric: $X \sim \mathrm{N}(\mu, \sigma^2)$ with unknown σ^2
    $\Rightarrow$ composite
  - nonparametric: X has finite mean
    $\Rightarrow$ composite

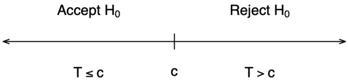
Hypothesis test

- Hypothesis test is a decision based on data
- The decision either accepts $\mathbb{H}_0$ or rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$
- Procedures of hypothesis testing
  - Construct a real valued function of the data called **test statistic**

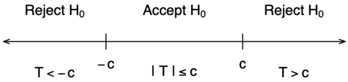
    $$T = T(X_1, X_2 \ldots X_n) \in \mathbb{R}$$

    which is a random variable
  - Pick a **critical region** C
    - One sided test: $C = \{x : x > c\}$ for **critical value** c
    - Two sided test: $C = \{x : |x| > c\}$ for **critical value** c
  - State hypothesis test as the decision rule

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } T \notin C \\ \text{reject } \mathbb{H}_0 & \text{if } T \in C \end{array}$$



(a) One-Sided Test



(b) Two-Sided Test

Figure: Acceptance and Rejection Regions for Test Statistic

Evaluation of hypothesis test

- A decision could be correct or incorrect
- We evaluate hypothesis tests through their probability of making mistakes
- Two types of errors in hypothesis testing

		Decision	
		Accept H_0	Reject H_0
Truth	$\mathbb{H}_0$	Correct decision	**Type I Error**
	$\mathbb{H}_1$	**Type II Error**	Correct decision

Power function

- Power function characterizes probability of making mistakes
- **Definition**: The **power function** of a hypothesis test is the probability of rejection

$$\pi(F) = P\{\text{reject } \mathbb{H}_0 | F\} = P\{T \in C | F\}$$

- **Definition**: The **size** of a hypothesis test is the probability of a Type I error

$$P\{\text{reject } \mathbb{H}_0 | F_0\} = \pi(F_0)$$

  for F_0 satisfying $\mathbb{H}_0$

- **Definition**: The **power** of a hypothesis test is the complement of the probability of a Type II error

$$P\{\text{reject } \mathbb{H}_0 | F_1\} = \pi(F_1) = 1 - P\{\text{accept } \mathbb{H}_0 | \mathbb{H}_1\}$$

  for F_1 satisfying $\mathbb{H}_1$

- Size is power function evaluated at null; Power is power function evaluated at alternative

Type I and II errors can't be reduced simultaneously

- Let $G(x|F) = P\{T \leq x|F\}$ be the sampling distribution of T
  - $G(x|F_0)$ is called **null sampling distribution**
  - $G(x|F_1)$ is called **alternative sampling distribution**
- Consider a one sided test with rejection rule $T > c$
  - Type I error is size $\pi(F_0) = P\{T > c|F_0\} = 1 - G(c|F_0)$
  - Type II error is $1 - \pi(F_1) = P\{T \leq c|F_1\} = G(c|F_1)$
- Since any distribution function $G(x|F)$ is increasing in x
  - Type I error is decreasing in c
  - Type II error is increasing in c

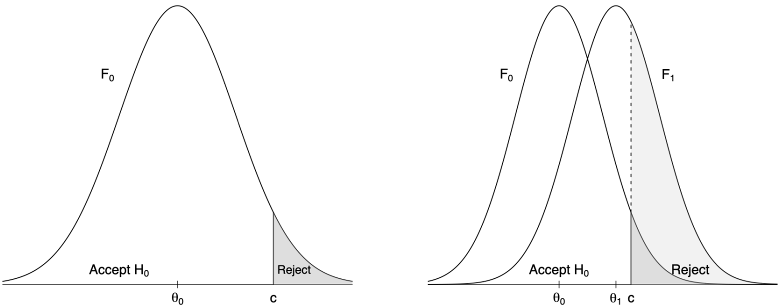


Figure: Left: Null Sampling Distribution for One-Sided Test; Right: Alternative Sampling Distribution for One-Sided Test

2. Classical Approach

Classical approach

- **Control size** and then pick the test to maximize the power subject to this size constraint
- **Definition**: The **significance level** $\alpha \in (0, 1)$ is the probability selected by the researcher to be the maximal acceptable size of the hypothesis test

Classical approach for one sided test

- Consider one sided test

$$\mathbb{H}_0 : \theta = \theta_0,\ \mathbb{H}_1 : \theta > \theta_0$$

- Given test statistic T, consider the test taking form

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } T \le c \\ \text{reject } \mathbb{H}_0 & \text{if } T > c \end{array}$$

- Choose c to control size at α

$$\pi(F_0) = P\{T > c | F_0\} = 1 - G(c|F_0) = \alpha \qquad (1)$$

- Solving (1) yields

$$c = G^{-1}(1 - \alpha | F_0),$$

the $(1 - \alpha)-$th quantile of the null sampling distribution

- The test rule

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } T \le G^{-1}(1 - \alpha | F_0) \\ \text{reject } \mathbb{H}_0 & \text{if } T > G^{-1}(1 - \alpha | F_0) \end{array}$$

has a size equal to α

Classical approach for two sided test

- Consider two sided test

$$\mathbb{H}_0 : \theta = \theta_0, \ \mathbb{H}_1 : \theta \neq \theta_0$$

with test taking form

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } |T| \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } |T| > c \end{array}$$

- Choose c to control size at α

$$\pi(F_0) = P\{|T| > c|F_0\} = 1 - G(c|F_0) + G(-c|F_0) = \alpha$$

- Suppose $G(x|F_0)$ is symmetric around 0

$$1 - G(c|F_0) + G(-c|F_0) = 2\left(1 - G(c|F_0)\right) = \alpha \qquad (2)$$

- Solving (2) yields

$$c = G^{-1}(1 - \frac{\alpha}{2}|F_0),$$

the $(1 - \frac{\alpha}{2})-$th quantile of the null sampling distribution

- The test rule

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } |T| \leq G^{-1}(1 - \frac{\alpha}{2}|F_0) \\ \text{reject } \mathbb{H}_0 & \text{if } |T| > G^{-1}(1 - \frac{\alpha}{2}|F_0) \end{array}$$

has a size equal to α

Example: T Test with normal sampling

- Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ and we wish to test

$$\mathbb{H}_0 : \mu = \mu_0, \ \mathbb{H}_1 : \mu > \mu_0$$

- Form test statistic

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

where $\bar{X}_n$ is sample mean and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

- Under $\mathbb{H}_0$

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{n-1}$$

- Given α, set

$$c = q_{1-\alpha}$$

where $q_{1-\alpha}$ is the $1-\alpha$-th quantile of t_{n-1} distribution

- A one sided t test with size α is

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } T \leq q_{1-\alpha} \\ \text{reject } \mathbb{H}_0 & \text{if } T > q_{1-\alpha} \end{array}$$

- If σ^2 is **known**, replacing s^2 with σ^2

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

  yields a z test that uses the quantile of a standard normal

- Analysis of a two sided test is similar

- **Theorem**: In the normal sampling model $X \sim \mathrm{N}(\mu, \sigma^2)$, let

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

1. The t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu > \mu_0$ rejects if

$$T > q_{1-\alpha}$$

   where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the t_{n-1} distribution

2. The t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu < \mu_0$ rejects if

$$T < q_{\alpha}$$

3. The t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu \neq \mu_0$ rejects if

$$|T| > q_{1-\alpha/2}$$

These tests have exact size α

Example: Asymptotic T test

- Again suppose X has mean μ and finite variance
- We wish to test

$$\mathbb{H}_0 : \mu = \mu_0, \ \mathbb{H}_1 : \mu > \mu_0$$

- The t-statistic is

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

where s^2 could be replaced by $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

- Under $\mathbb{H}_0$, T is not exactly normal but asymptotically normal by CLT

$$T \xrightarrow{d} \mathrm{N}(0, 1)$$

- Thus as $n \to \infty$

$$\pi(F_0) = P\{T > c | F_0\} \to P\{\mathrm{N}(0, 1) > c\} = 1 - \Phi(c)$$

- **Theorem**: If X has finite mean μ and variance σ^2

  1. The **asymptotic** t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu > \mu_0$ rejects if
  $$T > Z_{1-\alpha}$$
  where $Z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution

  2. The **asymptotic** t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu < \mu_0$ rejects if
  $$T < Z_{\alpha}$$

  3. The **asymptotic** t test of $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu \neq \mu_0$ rejects if
  $$|T| > Z_{1-\alpha/2}$$

  These tests have **asymptotic** size α

P-value

- Again consider a one sided test

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } T \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } T > c \end{array}$$

  where c is chosen to control size at α

$$P\{T > c|F_0\} = 1 - G(c|F_0) = \alpha$$

- How should we report the results of the test?
  - Method 1: report size α, and decision "Reject H_0" or "Accept H_0"
  - Method 2: report critical value c and value T at sample points
- Another method: report the value of a certain kind of statistic called **p-value**

- Define **p-value** as

$$p = 1 - G(T|F_0)$$

- Since $G(\cdot|F_0)$ is increasing, p is a decreasing function of T
- Also note

$$\alpha = 1 - G(c|F_0)$$

- Therefore, the decision

$$\text{reject } \mathbb{H}_0 \text{ if } T > c$$

  is equivalent to

$$\text{reject } \mathbb{H}_0 \text{ if } p < \alpha$$

Method 3: report the value of p

- For each $\alpha \in (0, 1)$

$$\begin{array}{ll} \text{accept } \mathbb{H}_0 & \text{if } p > \alpha \\ \text{reject } \mathbb{H}_0 & \text{if } p \leq \alpha \end{array}$$

  is a size α test

$$\begin{aligned} P\{p \leq \alpha | F_0\} &= P\{1 - G(T|F_0) \leq \alpha | F_0\} \\ &= P\{G^{-1}(1 - \alpha | F_0) \leq T | F_0\} \\ &= 1 - G\left(G^{-1}(1 - \alpha | F_0) | F_0\right) \\ &= \alpha \end{aligned}$$

- p is "degree of evidence against $\mathbb{H}_0$"
  - the smaller the p-value, the stronger the evidence against the null
- p is "marginal significance level"
  - the lower bound of the range of size α at which we would reject the null

Further remarks about p-value

- p is a transformation of a statistic rather than a probability
  - It transforms the T statistic to an easily interpretable universal scale between $[0, 1]$
- p allows inference to be continuous rather than dichotomous (more informative)
  - Suppose one statistic has p-value of 0.049 (mildly significant) and the second statistic has the p-value 0.051(mildly insignificant)
  - From their p value we know these two statistics are essentially the same
  - Reporting "Reject" or "Accept" would not be able to give us such information

2. Power Analysis

Introduction

- So far we focus on the size of the tests
- We know how to construct a test of (asymptotic) size α for mean
- A good test should also have a good power
- It is important to know the power of the test we constructed

Power of T test with known σ^2

- Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ with known σ^2
- Consider statistic

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

for tests

$$\mathbb{H}_0 : \mu = \mu_0, \;\; \mathbb{H}_1 : \mu > \mu_0$$

- We reject if

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > c$$

where c is chosen to control size at level α

- Whether $\mathbb{H}_0$ is true or not, $\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathrm{N}(0,1)$ since $\bar{X}_n$ is centered around true mean μ
- The power function of the test is

$$\begin{aligned}
\pi(F) = P\{T > c|F\} &= P\left\{\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > c|F\right\} \\
&= P\left\{\underbrace{\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}}_{Z \sim \mathrm{N}(0,1)} + \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > c|F\right\} \\
&= 1 - \Phi\left(c + \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right)
\end{aligned}$$

- Size is $\pi(F_0) = 1 - \Phi(c)$, since $F_0 = \{F \sim \mathrm{N}(\mu, \sigma^2) : \mu = \mu_0\}$
- Power is $\pi(\mu|F_1) = 1 - \Phi\left(c + \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right)$ where $\mu > \mu_0$
  - Note $\pi(\mu|F_1)$ is increasing in n, μ and decreasing in σ^2 and c

Example: Selection of c and n for power targets

- Suppose now we want to select n and c to achieve size 0.1 and

  power at least 0.8 if $\mu \geq \mu_0 + \sigma$

- How should we proceed?
- Step 1: selecting c such that

$$\pi(F_0) = 1 - \Phi(c) = 0.1 \tag{3}$$

  ensures size $\alpha = 0.1$. Solving (3) yields $c = 1.28$

- Step 2: since power is increasing in μ, selecting n such that

$$1 - \Phi\left(1.28 + \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} | \mu = \mu_0 + \sigma\right) \geq 0.8$$

  Solving above inequality yields $n \geq 4.49$

- Conclusion: choosing $c = 1.28$ and $n = 5$ yields the desired size and power balance

3. Likelihood Ratio Test

Motivation

- Recall classical approach to testing
  - **Control size** and then pick the test to maximize power subject to this size constraint
- So far we focus on t test
- Another important class of tests is **likelihood ratio test**
  - We show it maximizes power subject to size constraint for testing **simple hypotheses**

Likelihood ratio test for simple hypotheses

- Consider a parametric model $f(x|\theta)$ with likelihood $L_n(\theta) = \prod_{i=1}^n f(X_i|\theta)$

- We want to test simple hypotheses

$$\mathbb{H}_0 : \theta = \theta_0, \ \mathbb{H}_1 : \theta = \theta_1$$

  for some hypothetical values θ_0 and θ_1

- The ratio $\frac{L_n(\theta_1)}{L_n(\theta_0)}$ compares the likelihood function under two hypotheses

- A decision rule could be

$$\begin{array}{ll} \text{accept } \mathbb{H}_0 & \text{if } \frac{L_n(\theta_1)}{L_n(\theta_0)} \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } \frac{L_n(\theta_1)}{L_n(\theta_0)} > c \end{array}$$

  for some critical value c

- For convenience, define the **likelihood ratio statistic** as

$$LR_n = 2\left(\ell_n(\theta_1) - \ell_n(\theta_0)\right)$$

where $\ell_n(\theta) = \log L_n(\theta)$

- A likelihood ratio test is

$$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } LR_n \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } LR_n > c \end{array}$$

for some critical value c

Example: normal sampling with known variance

- For $X \sim \mathrm{N}(\mu, \sigma^2)$ with known σ^2

$$\ell_n(\mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

- Suppose

$$\mathbb{H}_0 : \mu = \mu_0, \; \mathbb{H}_1 : \mu = \mu_1 > \mu_0$$

$$\begin{aligned} LR_n &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left((X_i - \mu_0)^2 - (X_i - \mu_1)^2\right) \\ &= \frac{n}{\sigma^2}\left[2\bar{X}_n(\mu_1 - \mu_0) + (\mu_0^2 - \mu_1^2)\right] \end{aligned}$$

- Rejecting $\mathbb{H}_0$ for some $LR_n > c$ is equivalent to rejecting if

$$T = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \text{some constant}$$

Neyman-Pearson Lemma

- **Theorem**: Suppose random variable X has a parametric pdf/pmf $f(X|\theta)$. Among all tests of a simple null hypothesis against a simple alternative hypothesis

$$\mathbb{H}_0 : \theta = \theta_0, \ \mathbb{H}_1 : \theta = \theta_1$$

  with size α, the likelihood ratio test has the greatest power.

- In the normal sampling model with known variance, the likelihood ratio test of simple hypotheses is identical to a t test using a known variance

- By Neyman-Pearson Lemma, t test using a known variance is the most powerful test for this hypothesis in this model

Proof

- Consider likelihood ratio test

$$\begin{array}{ll} \text{accept } \mathbb{H}_0 & \text{if } \frac{L_n(\theta_1)}{L_n(\theta_0)} \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } \frac{L_n(\theta_1)}{L_n(\theta_0)} > c \end{array}$$

  where c is chosen such that

$$P\left\{\frac{L_n(\theta_1)}{L_n(\theta_0)} > c|\theta = \theta_0\right\} = \alpha$$

- Let the joint density of observations be $f(\mathbf{x}|\theta)$ for some $\mathbf{x} = (x_1, \dots x_n)'$
- Then $L_n(\theta) = f(\mathbf{X}|\theta)$, where $\mathbf{X} = (X_1, \dots X_n)$

- Since test is binary decision (accept/reject), it can be represented by binary function (called test function)
- The likelihood ratio test function is

$$\psi_{LR} = \mathbf{1}\left\{f(\mathbf{X}|\theta_1) > cf(\mathbf{X}|\theta_0)\right\}$$

that is, $\psi_{LR} = 1$ if likelihood ratio rejects $\mathbb{H}_0$ and $\psi_{LR} = 0$ otherwise

- Let ψ_a be any alternative test function with same size α
- Since both tests have same size

$$P\left\{\psi_{LR} = 1|\theta = \theta_0\right\} = P\left\{\psi_a = 1|\theta = \theta_0\right\} = \alpha$$

or equivalently

$$\int \psi_{LR} f(\mathbf{x}|\theta_0)d\mathbf{x} = \int \psi_a f(\mathbf{x}|\theta_0)d\mathbf{x} = \alpha$$

- The power of likelihood ratio test is

$$
\begin{aligned}
& P\left\{\frac{L_n(\theta_1)}{L_n(\theta_0)} > c|\theta = \theta_1\right\} \\
=& P\left\{\psi_{LR} = 1|\theta = \theta_1\right\} \\
=& \int \psi_{LR} f(\mathbf{x}|\theta_1)dx \\
=& \int \psi_{LR} f(\mathbf{x}|\theta_1)dx - c\left\{\int \psi_{LR} f(\mathbf{x}|\theta_0)d\mathbf{x} - \int \psi_a f(\mathbf{x}|\theta_0)d\mathbf{x}\right\} \\
=& \int \psi_{LR}\left(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)\right)dx + c\int \psi_a f(\mathbf{x}|\theta_0)d\mathbf{x} \\
\geq& \int \psi_a\left(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)\right)dx + c\int \psi_a f(\mathbf{x}|\theta_0)d\mathbf{x} \\
=& \int \psi_a f(\mathbf{x}|\theta_1)dx \\
=& \text{power of } \psi_a
\end{aligned}
$$

- The inequality holds since
  - if $(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)) > 0$, $\psi_{LR} = 1$, and

$$\psi_{LR}\,(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)) \geq \psi_a\,(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0))$$

  - if $(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)) \leq 0$, $\psi_{LR} = 0$

$$\psi_{LR}\,(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0)) = 0 \geq \psi_a\,(f(\mathbf{x}|\theta_1) - cf(\mathbf{x}|\theta_0))$$

- Hence the power of the likelihood ratio test is greater than the power of the test ψ_a
- By the arbitrariness of ψ_a, we conclude likelihood ratio test has higher power than any other test with the same size

Likelihood Ratio Test against composite alternatives

- Consider **two sided test**

$$\mathbb{H}_0 : \theta = \theta_0,\ \mathbb{H}_1 : \theta \neq \theta_0$$

- The log likelihood under $\mathbb{H}_1$ is the unrestricted maximum of the likelihood
- Let $\hat{\theta}$ be the MLE that maximizes $L_n(\theta)$
- The likelihood ratio statistic is

$$LR_n = 2\left(\ell_n(\hat{\theta}) - \ell_n(\theta_0)\right)$$

- The likelihood ratio test is

$$\begin{array}{ll} \text{accept } \mathbb{H}_0 & \text{if } LR_n \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } LR_n > c \end{array}$$

for some critical value c

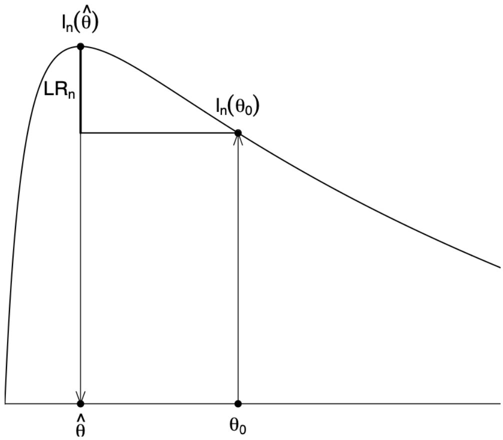


Figure: Likelihood Ratio

- Consider **one sided test**

$$\mathbb{H}_0 : \theta = \theta_0, \ \mathbb{H}_1 : \theta > \theta_0$$

- The log likelihood under $\mathbb{H}_1$ is the maximum of the log likelihood in the restricted set

$$\{\theta : \theta \geq \theta_0\},$$

that is, $\ell_n(\hat{\theta}^+)$, where $\hat{\theta}^+ = \arg\max_{\theta \geq \theta_0} \ell_n(\theta)$

- The likelihood ratio statistic is

$$LR_n^+ = 2\left(\ell_n(\hat{\theta}^+) - \ell_n(\theta_0)\right)$$

- The likelihood ratio test is

$$\begin{array}{cl} \text{accept } \mathbb{H}_0 & \text{if } LR_n^+ \leq c \\ \text{reject } \mathbb{H}_0 & \text{if } LR_n^+ > c \end{array}$$

for some critical value c

Example: Normal sampling with known variance

- Again suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ with σ^2 known
- Consider testing

$$\mathbb{H}_0 : \mu = \mu_0,\ \mathbb{H}_1 : \mu > \mu_0$$

- We've shown that for simple hypothesis

$$\mathbb{H}_0 : \mu = \mu_0,\ \mathbb{H}_1 : \mu = \mu_1 > \mu_0$$

likelihood ratio test is equivalent to a t test

$$\text{rejecting } H_0 \text{ if } \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} > b, \text{ for some } b$$

- Such analysis does not depend on specific value of μ_1
- Thus this t test is also the likelihood ratio test for one-sided alternative

Asymptotic size control for Likelihood Ratio Test

- **Theorem**: For simple null hypotheses, under $\mathbb{H}_0 : \theta = \theta_0$

  $$LR_n \xrightarrow{d} \chi^2_{\dim(\theta)}$$

  Let $q_{1-\alpha}$ be the $1-\alpha$-th quantile of $\chi^2_{\dim(\theta)}$. The test

  $$\begin{array}{rl} \text{accept } \mathbb{H}_0 & \text{if } LR_n \leq q_{1-\alpha} \\ \text{reject } \mathbb{H}_0 & \text{if } LR_n > q_{1-\alpha} \end{array}$$

  has asymptotic size α

- Moreover, likelihood ratio and t tests are **asymptotically** equivalent tests

Sketch proof

- Note $LR_n = 2\left(\ell_n(\hat{\theta}) - \ell_n(\theta_0)\right)$
- Second order Taylor expansion yields

$$\ell_n(\theta_0) \simeq \ell_n(\hat{\theta}) + \underbrace{\frac{\partial}{\partial\theta}\ell_n(\hat{\theta})}_{\mathbf{0}}'(\hat{\theta}-\theta_0) + \frac{1}{2}(\hat{\theta}-\theta_0)'\underbrace{\frac{\partial^2}{\partial\theta\partial\theta}\ell_n(\hat{\theta})}_{-\hat{V}^{-1}}(\hat{\theta}-\theta_0)$$

- Note where $\hat{V} = \left\{-\frac{\partial^2}{\partial\theta\partial\theta}\ell_n(\hat{\theta})\right\}^{-1}$ is the Hessian estimator of the asymptotic variance of $\hat{\theta}$ estimated Hessian
- Hence

$$2\left(\ell_n(\hat{\theta}) - \ell_n(\theta_0)\right) \simeq (\hat{\theta}-\theta_0)'\hat{V}^{-1}(\hat{\theta}-\theta_0)$$

- As $n \to \infty$, the RHS converges to $\chi^2_{\dim(\theta)}$