Lecture 4: Hypothesis Testing

Econ 205A: Econometric Methods I

Ruoyao Shi

Fall 2018

Contents

1 Methods of Finding Tests

Basic Concepts on Hypothesis Testing

- Suppose we have a statistical model and data $X_1, \ldots, X_n$ whose distribution depends on a (scalar or vector) parameter θ, which specifies the true state of nature. Assume θ ranges over a known parameter space Θ. In general, we have two competing hypotheses about the parameter θ of the model.

- **Definition 1** *A **hypothesis** is a statement about the population parameter; or equivalently, a statement about the joint distribution of the random variables in the model.*

- **Definition 2** *The two competing hypotheses are called **null hypothesis** and **alternative hypothesis**, respectively. The null hypothesis* *is denoted as H_0, and the alternative hypothesis is denoted as H_1 (or sometimes H_a).*

- The general form of the null hypothesis is

$$H_0 : \theta \in \Theta_0,$$
$$H_1 : \theta \in \Theta_1 \equiv \Theta_0^c = \Theta \backslash \Theta_0;$$

where Θ_0 is a subset of the parameter space Θ.

- Note that $\Theta_0 \cap \Theta_1 = \emptyset$, and $\Theta_0 \cup \Theta_1 = \Theta$.

- **Example 1** *Suppose that $X_1, \ldots, X_{10}$ are from the distribution $\mathcal{N}(\mu, 9)$, where μ is unknown.*

$$H_0 : \mu = 0,$$
$$H_1 : \mu = 1.$$

In this example, both the null and the alternative are simple hypotheses (defined below).

- There are two types of hypotheses:
  - **Simple hypothesis** is a hypothesis that consists of only a single joint distribution of the random variables in the model.
  - **Composite hypothesis** is a hypothesis that consists of more than one joint distribution.

- **Example 2** *Suppose that $X_1, \ldots, X_{10}$ are from the distribution $\mathcal{N}(\mu, 9)$, where μ is unknown.*
  - *Suppose $\Theta = \{(\mu, 9) : \mu \geq 0\}$. Then $H_0 : \mu = 0$ v.s. $H_1 : \mu > 0$ is simple v.s. composite hypotheses.*
  - *Suppose $\Theta = \{(\mu, 9) : \mu \in \mathbb{R}\}$. Then $H_0 : \mu = 4$ v.s. $H_1 : \mu \neq 4$ is also simple v.s. composite hypotheses.*
  - *Suppose $\Theta = \{(\mu, 9) : \mu \in \mathbb{R}\}$. Then $H_0 : \mu \leq -2$ v.s. $H_1 : \mu > -2$ is also composite v.s. composite hypotheses.*

- **Example 3** *Suppose that* $X_1, \ldots, X_{10}$ *are from the distribution* $\mathcal{N}(\mu, \sigma^2)$*, and* $\Theta = \{(\mu, \sigma^2) : \sigma^2 > 0, \text{ and } \mu = -1 \text{ or } 3\}$*. Then* $H_0 : \mu = -1$ *v.s.* $H_1 : \mu = 3$ *is composite v.s. composite hypotheses. Even though both* H_0 *and* H_1 *specify only one value for* μ*, because* σ^2 *is unknown, they do not pin down the distribution for the sample.*

- The goal of a hypothesis test is to decide, based on the data, which of the two competing hypotheses is true. The null hypothesis is the maintained hypothesis that is held to be true until sufficient evidence indicates the opposite. The alternative hypothesis is the hypothesis against which the null is tested.

- **Definition 3** *A* ***hypothesis testing procedure*** *or* ***hypothesis test*** *is a rule that specifies:*

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

  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 critical region is often denoted as* C*.*

- If the sample values fall into the critical region, then the evidence is in favor of H_1; if in the acceptance region, then the evidence is in favor of H_0. Characterization of the critical region is equivalent to characterization of a test.

- **Example 4** *Suppose we have a sample* X_1, X_2 *from the distribution* $\mathcal{N}(\mu, 1)$*, where the unknown mean* $\mu \in \mathbb{R}$*. We are interested in testing* $H_0 : \mu = 0$ *v.s.* $H_1 : \mu \neq 0$*. A commonly used test is the t test, which rejects* H_0 *if*

  $$\left| \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} \right| = \left| \frac{\sqrt{2}\bar{X}_2}{1} \right| > 1.96,$$

  which is equivalent to $|\bar{X}_2| > 1.96/\sqrt{2}$*. (Draw the critical region* $\{(X_1, X_2) : |\bar{X}_2| > 1.96/\sqrt{2}\} = \{(X_1, X_2) : X_1 + X_2 > 1.96\sqrt{2} \text{ or } X_1 + X_2 < -1.96\sqrt{2}\}$ *in class.)*

- Typically, a hypothesis test is characterized in terms of a **test statistic** and a **critical value**. In the above example, $|\bar{X}_2|$ is the test statistic, $1.96/\sqrt{2}$ is the critical value.

- Usually, we want the (asymptotic) distribution of the test statistic under H_0 to be known.

Likelihood Ratio (LR) Test

- Likelihood ratio test is related to the maximum likelihood estimator, and is widely used.

- Let $X_1, \ldots, X_n$ be a random sample from a population with pdf or pmf $f(x|\theta)$, the likelihood function is

$$L(\theta|x_1, \ldots, x_n) = L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \Pi_{i=1}^n f(x_i|\theta).$$

- **Definition 4** *The* ***likelihood ratio test statistic*** *for testing* H_0 : $\theta \in \Theta_0$ *versus* $H_1 : \theta \in \Theta_0^c$ *is*

$$\lambda(\mathbf{x}) \equiv \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{x})}.$$

- Note that the infimum in the denominator is taken over the entire parameter space Θ, not Θ_1. Since $\Theta_0 \subset \Theta$, the numerator is no larger than the denominator.

- **Definition 5** *A* ***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* $0 \leq c \leq 1$.

- The value of c depends on the distribution of $\lambda(\mathbf{x})$ in each model and needs to be discussed case by case.

- Recall that the likelihood is the probability that the observed sample would have been generated if the true parameter value were to be θ. Then the numerator of $\lambda(\mathbf{x})$ is the supreme probability that the observed sample would have been generated if H_0 were true; and the denominator is the supreme probability that the observed sample would have been generated if $\theta \in \Theta$ at all. The fact that the numerator is very small serves as the evidence against H_0. If the numerator is small enough (relative to the deonominator), we should reject H_0.

- **Example 5** *Suppose on a Saturday evening, you and your roommate finish all the first year econ PhD course work and decide to roll a die* 100 *times to kill time. You get* 99 *times of number* 3 *and once other numbers. Do you think the die is fair? Of course not! Because you know that if the die was fair, it is almost impossible to have obtained such outcome. Likelihood ratio test is the formalization of such intuition. Formally, let p denote the probability that the die shows number* 3*, then the hypotheses we want to test are:*

$$H_0 : p = \frac{1}{6} \text{ v.s. } H_1 : p \neq \frac{1}{6}.$$

Under H_0, the probability of the observed sample (the likelihood) is $\left(\frac{1}{6}\right)^{99}\left(\frac{5}{6}\right)^{1} = 7.65 \times 10^{-78}$. *Note that the MLE of p is* $99/100 = 0.99$*, then the probability of the observed sample is* $0.99^{99}0.01^{1} = 0.0037$. *Among all the possible values of p, $p = 1/6$, the one indicated by H_0 is too unlikely given the sample outcome, so we should reject H_0.*

- **Example 6** *Let* $\mathbf{X} = (X_1, \ldots, X_n)'$ *be a random sample from* $\mathcal{N}(\mu, \sigma^2)$ *with σ^2 known ($\sigma^2 = \sigma_0^2$). Let*

$$H_0 : \mu = \mu_0 \text{ v.s. } H_1 : \mu \neq \mu_0.$$

Then

$$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\},$$
$$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 = \sigma_0^2\}.$$

Let $\hat{\mu} = \bar{X}_n$ denote the MLE of μ, then the likelihood ratio is

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})} \\
&= \frac{(2\pi\sigma_0^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i - \mu_0)^2\right)}{(2\pi\sigma_0^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i - \hat{\mu})^2\right)} \\
&= \exp\left(-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (x_i - \mu_0)^2 - (x_i - \hat{\mu})^2\right]\right) \\
&= \exp\left(-\frac{n}{2\sigma_0^2}(\hat{\mu} - \mu_0)^2\right).
\end{aligned}$$

This implies that

$$-2\log\lambda(\mathbf{X}) = \frac{n}{\sigma_0^2}(\hat{\mu}-\mu_0)^2 = \left(\frac{\bar{X}_n-\mu_0}{\sigma_0/\sqrt{n}}\right)^2 \sim \chi_1^2 \text{ under } H_0.$$

Therefore, at the significance level α,

$$P(-2\log\lambda(\mathbf{X}) \geq \chi^2_{1,1-\alpha}) = P(\lambda(\mathbf{X}) \leq \exp(-\chi^2_{1,1-\alpha}/2) = \alpha;$$

that is, the critical value c *should be* $\exp(-\chi^2_{1,1-\alpha}/2)$ *in this example.*

- **Example 7** *Now suppose* σ^2 *is unknown, and still we want to test*

$$H_0 : \mu = \mu_0 \text{ v.s. } H_1 : \mu \neq \mu_0.$$

It is not hard to show that

$$\lambda(\mathbf{x}) = \left(\frac{n^{-1}\sum_{i=1}^n (x_i-\mu_0)^2}{n^{-1}\sum_{i=1}^n (x_i-\bar{x})^2}\right)^{-n/2}. \tag{1}$$

Since

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

we have

$$\begin{aligned}
\lambda(\mathbf{x})^{-2/n} &= 1 + \frac{(\bar{x}-\mu_0)^2}{n^{-1}\sum_{i=1}^n (x_i-\bar{x})^2} \\
\Rightarrow T_n \equiv \left(\lambda(\mathbf{x})^{-2/n} - 1\right) / \left(\frac{1}{n-1}\right) &= \frac{(\bar{x}-\mu_0)^2/n}{(n-1)^{-1}\sum_{i=1}^n (x_i-\bar{x})^2} \\
&= \left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}}\right)^2 \\
&\sim t^2_{n-1} \\
&\overset{d.}{=} F_{1,n-1} \text{ under } H_0.
\end{aligned}$$

Therefore, at the significance level α,

$$P(T_n \geq F_{1,n-1,1-\alpha}) = P\left(\lambda(\mathbf{X}) \leq \left(\frac{F_{1,n-1,1-\alpha}}{n-1}+1\right)^{-n/2}\right) = \alpha;$$

that is, the critical value c equals $\left(\frac{F_{1,n-1,1-\alpha}}{n-1}+1\right)^{-n/2}$ in this example. Recall that $F_{1,n-1} \xrightarrow{d.} \chi^2_1$, so the critical could be approximated by $\left(\frac{\chi^2_{1,1-\alpha}}{n-1}+1\right)^{-n/2} \approx \exp(-\chi^2_{1,1-\alpha}/2)$ as $n \to \infty$.

Note that in the above example, the LR test is equivalent to a test based on the Student's t statistic.

2 Methods of Evaluating Tests

Error Probabilities and the Power Function

- The outcome of the test depends on realization of the random variables in the sample, and hence errors can be made. Two types of errors are possible:
  - **Type I error**: reject H_0 when H_0 is true;
  - **Type II error**: accept H_0 when H_0 is false.
- The probability that a test rejects H_0 depends on the value of the parameter θ. It is called the **power function** of the test; that is

$$\beta_n(\theta) \equiv P_\theta(\text{test rejects } H_0).$$

- Note that $\beta_n(\theta)$ usually also depends on n. $\beta(\theta) \equiv \lim_{n\to\infty} \beta_n(\theta)$ is called the **asymptotic power function**.
- The *probability of type I error*, $P_\theta(\text{test rejects } H_0)$ when $\theta \in \Theta_0$, is the probability that the test rejects H_0 when the true parameter value $\theta \in \Theta_0$. In other words, the probability of type I error is the probability that the sample falls into the critical region when the true parameter value $\theta \in \Theta_0$. Again, the probability of type I error depends on the value of θ.
- When H_0 is a composite hypothesis, the maximum probability of type I error over the null parameter space is called the **size** of the test; that is

$$\textbf{size} \text{ of the test } = \sup_{\theta\in\Theta_0} P_\theta(\text{test rejects } H_0).$$

- **Definition 6** *We say that a test has **significance level** α ($\alpha \in (0,1)$) if its size is less than or equal to α.*

- It is convention to consider significance levels of 0.01, 0.05 or 0.10.

- Note that if a test ψ has significance level α_1, and $\alpha_1 < \alpha_2$, then ψ must also has significance level α_2.

- The **power** of the test is the probability of rejecting H_0 when H_1 is true; equivalently, the power equals one minus the probability of type II error. That is,

  $$\textbf{power} \text{ of the test } = P_{\theta \in \Theta_1}(\text{test rejects } H_0).$$

  The probability of type II error is often denoted by β. Then the power of the test is denoted by $1 - \beta$.

- Probabilities of type I and type II errors, size and power are inherent properties of a test.

- **Definition 7** *The smallest value α for which the given data rejects the null hypothesis H_0 at the significance level α is called the **p-value**. In other words, p-value is the infimum of the significance level α at which the given data rejects the null hypothesis.*

- The p-value depends on a particular sample. The importance of the p-value is that once the p-value is calculated, we can easily decide whether or not to reject H_0 at any significance level α:

  - If $p < \alpha$, reject H_0 at significance level α;
  - If $p \geq \alpha$, accept H_0 at significance level α.

- Trade-off between size and power.

- **Definition 8** *A test is called **consistent** if the asymptotic power function $\beta(\theta)$ equals to one for all $\theta \in \Theta_1$.*

Most Powerful Tests

- Recall that characterization of a test is equivalent to characterization of the critical region. The larger the critical region is, the larger the probability to reject H_0, whether $\theta \in \Theta_0$ or $\theta \in \Theta_1$. That is, the larger the size (probability of type I error) and the larger the power (one minus probability of type II error). Thus, there is in general a trade-off between probability of type I and type II errors. They cannot be minimized at the same time.

- The conventional approach of hypothesis testing is to first control the size α (for choosen α), and then find the most powerful test among all the size α tests.

- **Definition 9** *A test is called **uniformly most powerful (UMP)** test of size α if it has higher power than any other tests of size α for all $\theta \in \Theta_1$.*

- **Theorem 1** ***(Neyman-Pearson Lemma)*** *Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf or pmf corresponding to θ_i is $f(\mathbf{x}|\theta_i)$, $i = 0, 1$. Consider the test based on the following critical region*
$$C \equiv \left\{ x : \frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} = \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} \geq k_\alpha \right\},$$
where k_α is chose such that the significance level is α for given α. Then C is the critical region of a UMP test of size α.

- The Neyman-Pearson lemma suggests a method of constructing UMP tests based on the likelihood ratio.

- The intuition is the same as the likelihood ratio test: if the likelihood $L(\theta_1|\mathbf{x})$ is large relative to $L(\theta_0|\mathbf{x})$, then the sample is more likely to have been generated with the parameter value θ_1 instead of θ_0, which leads to the rejection of H_0.

- *Proof of Theorem 1.* Assume that A is the critical region of another test at significance level α. Then we want to show that
$$\int_C f(\mathbf{x}|\theta_1)d\mathbf{x} \geq \int_A f(\mathbf{x}|\theta_1)d\mathbf{x}.$$

Using the indicator function, the above inequality can be written as

$$\int (I_C(\mathbf{x}) - I_A(\mathbf{x}))f(\mathbf{x}|\theta_1)d\mathbf{x} \geq 0.$$

To prove this inequality, first assume that we already proved

$$(I_C(\mathbf{x}) - I_A(\mathbf{x}))f(\mathbf{x}|\theta_1) \geq k_\alpha(I_C(\mathbf{x}) - I_A(\mathbf{x}))f(\mathbf{x}|\theta_0). \tag{2}$$

We assume so because if (2) holds, then

$$\begin{aligned}\int (I_C(\mathbf{x}) - I_A(\mathbf{x}))f(\mathbf{x}|\theta_1)d\mathbf{x} &\geq k_\alpha \int (I_C(\mathbf{x}) - I_A(\mathbf{x}))f(\mathbf{x}|\theta_0)d\mathbf{x} \\ &= k_\alpha \int_C f(\mathbf{x}|\theta_0)d\mathbf{x} - k_\alpha \int_A f(\mathbf{x}|\theta_0)d\mathbf{x} \\ &= k_\alpha\alpha - k_\alpha\alpha \\ &= 0,\end{aligned}$$

which is the conclusion of the theorem. Now we prove (2). Since I_C and I_A are both indicator functions, $I_C(\mathbf{x}) - I_A(\mathbf{x})$ can only take three possible values: 1, 0 or -1. When $I_C(\mathbf{x}) - I_A(\mathbf{x}) = 0$, we trivially have (2). When $I_C(\mathbf{x}) - I_A(\mathbf{x}) = 1$, then we must have

$$f(\mathbf{x}|\theta_1) \geq k_\alpha f(\mathbf{x}|\theta_0),$$

which implies (2). When $I_C(\mathbf{x}) - I_A(\mathbf{x}) = -1$, then we must have

$$f(\mathbf{x}|\theta_1) < k_\alpha f(\mathbf{x}|\theta_0),$$

which again implies 2). Thus we can conclude that the inequality (2) holds and this completes the proof.

- **Example 8** ***(Binomial UMP Test)*** *Let $X \sim binomial(2, \theta)$. We want to find the UMP test for $H_0 : \theta = \frac{1}{2}$ versus $H_1 : \theta = \frac{3}{4}$. Note that the likelihood function is*

$$L(\theta|x) = f(x|\theta) = \binom{2}{x} \theta^x(1-\theta)^{2-x}.$$

So the likelihood ratios for different possible samples are

$$\frac{L(\theta = \frac{3}{4}|x=0)}{L(\theta = \frac{1}{2}|x=0)} = \frac{1}{4},$$
$$\frac{L(\theta = \frac{3}{4}|x=1)}{L(\theta = \frac{1}{2}|x=1)} = \frac{3}{4},$$
$$\frac{L(\theta = \frac{3}{4}|x=2)}{L(\theta = \frac{1}{2}|x=2)} = \frac{9}{4}.$$

The Neyman-Pearson lemma tells us that the UMP test of level α should reject H_0 if the likelihood ratio is greater than or equal to k_α. If we choose $k > \frac{9}{4}$, we will never reject H_0, hence the size of the test will be zero. If we choose $\frac{3}{4} < k < \frac{9}{4}$, then we will reject H_0 when $x = 2$, so the size of the test will be

$$\alpha = P(X = 2|\theta = \frac{1}{2}) = \frac{1}{4}.$$

In other words, any likelihood ratior test with critical value $\frac{3}{4} < k_\alpha < \frac{9}{4}$ is a UMP test at level $\alpha = \frac{1}{4}$. Similarly, if we choose $\frac{1}{4} < k < \frac{3}{4}$, then we will reject H_0 when $x = 2$ or $x = 1$, so the size of the test will be

$$\alpha = 1 - P(X = 0|\theta = \frac{1}{2}) = \frac{3}{4}.$$

In other words, any likelihood ratiro test with critical value $\frac{1}{4} < k_\alpha < \frac{3}{4}$ is a UMP test at level $\alpha = \frac{3}{4}$. Finally, if we choose $k < \frac{1}{4}$, then we will always reject H_0, henc the size of the test will be one.

- Question: what if we specify $\alpha = 0.05$? We need to randomize the test: reject H_0 $\frac{1}{5}$ of the times we get $x = 2$, then the size of the test will be exactly $\alpha = \frac{1}{5}P(X = 2|\theta = \frac{1}{2}) = \frac{1}{5}\frac{1}{4} = 0.05$.

- **Example 9** ***(Normal UMP Test)*** *Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ population, and assume that σ^2 is known. We want to find the UMP test for the hypotheses*

$$H_0 : \mu = \mu_0 \textit{ versus } H_1 : \mu = \mu_1.$$

The Neyman-Pearson lemme tells us that the UMP test rejects H_0 if

$$\begin{aligned}\frac{L(\mu_1|\mathbf{X})}{L(\mu_0|\mathbf{X})} &= \frac{(1/\sqrt{2\pi\sigma^2})^{-n}\exp\left[-\sum_{i=1}^n (X_i-\mu_1)^2/2\sigma^2\right]}{(1/\sqrt{2\pi\sigma^2})^{-n}\exp\left[-\sum_{i=1}^n (X_i-\mu_0)^2/2\sigma^2\right]}\\ &= \exp\left[\left(\sum_{i=1}^n X_i\right)(\mu_1-\mu_0)/\sigma^2 - n(\mu_0^2-\mu_1^2)/2\sigma^2\right]\\ &\geq k_\alpha.\end{aligned}$$

This inequality is equivalent to

$$\begin{cases} \bar{X}_n \geq c_\alpha & \text{if } \mu_0 < \mu_1,\\ \bar{X}_n \leq c_\alpha & \text{if } \mu_0 > \mu_1,\end{cases}$$

where c_α is a constant related to k_α. To figure out the critical value c_α, note that

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

So when $\mu_0 < \mu_1$, $P_{\mu_0}(\bar{X}_n \geq \frac{\sigma Z_{1-\alpha}}{\sqrt{n}} + \mu_0) = \alpha$; that is $c_\alpha = \frac{\sigma Z_{1-\alpha}}{\sqrt{n}} + \mu_0$. Similarly, when $\mu_0 > \mu_1$, $c_\alpha = \frac{\sigma Z_\alpha}{\sqrt{n}} + \mu_0$.

- For simple versus simple hypotheses, the Neyman-Pearson lemma leads to the test statistic based on the likelihood ratio. But to determine the critical value k_α (or c_α after the transformation), we need to utilize the probability knowledge we have learned before.

- Note that μ_1 only determines the direction of the inequality for the critical value, but the critical value only depends on μ_0.

- **Example 10** ***(Exponential UMP Test)*** *Let $X_1, \ldots, X_n$ be a random sample from the exponential(θ) population. We want to find the UMP test for the hypotheses*

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta = \theta_1.$$

Using the Neyman-Pearson lemma, we can show that the UMP test rejects H_0 if

$$\begin{cases} \bar{X}_n \geq c_\alpha & \text{if } \theta_0 > \theta_1,\\ \bar{X}_n \leq c_\alpha & \text{if } \theta_0 < \theta_1.\end{cases} \tag{3}$$

To find the critical value c_α, we use the normal approximation (CLT). First note that $\mathbb{E}(X_i) = 1/\theta_0$ and $var(X_i) = 1/\theta_0^2$. When $\theta_0 < \theta_1$, by CLT we have that $\frac{\sqrt{n}(\bar{X}_n - 1/\theta_0)}{1/\theta_0} \approx \mathcal{N}(0,1)$, so the critical value is $c_\alpha = \frac{Z_\alpha}{\theta_0\sqrt{n}} + \frac{1}{\theta_0}$. Similarly, we can show that when $\theta_0 > \theta_1$, the critical value is $c_\alpha = \frac{Z_{1-\alpha}}{\theta_0\sqrt{n}} + \frac{1}{\theta_0}$.

- Note again that θ_1 only determines the direction of the inequality for the critical value, but the critical value only depends on θ_0.

Most Powerful Tests for Composite Hypotheses

- The Neyman-Pearson lemma states that the UMP test for simple versus simple hypotheses should be based on the likelihood ratios.
- Even though it does not directly suggest the UMP tests for composite hypotheses, in some cases, we can find the UMP test (or show the non-existence of the UMP test) with the help of the Neyman-Pearson lemma.
- In Examples 9 and 10, the UMP tests take the same form as long as the alternative parameter values are on the same side of the null parameter value. In addition, the alternative parameter value only affects the form of the critical region, but the critical value only depends on the null parameter value.
- By this argument, we get that the test found in Example 9 is also the UMP test for the following hypotheses:

$$H_0: \mu = \mu_0 \text{ v.s. } H_1: \mu > \mu_0; \text{ or}$$
$$H_0: \mu = \mu_0 \text{ v.s. } H_1: \mu < \mu_0.$$

And similarly, the test found in Example 10 is also the UMP test for the following hypotheses:

$$H_0: \theta = \theta_0 \text{ v.s. } H_1: \theta > \theta_0; \text{ or}$$
$$H_0: \theta = \theta_0 \text{ v.s. } H_1: \theta < \theta_0.$$

- For two-sided alternative hypothesis such as

$$H_0 : \mu = \mu_0 \text{ v.s. } H_1 : \mu \neq \mu_0,$$

  there does not exist a UMP test. The reason is that the UMP test for alternative parameter $\mu_1 > \mu_0$ is different from the UMP test for $\mu_1 < \mu_0$, so a single test cannot have highest power for all values $\mu_1 \neq \mu_0$.

- What about composite v.s. composite hypotheses? We modify Example 10 to illustrate the argument now. Suppose we are interested in finding the UMP test for the hypotheses

$$H_0 : \theta \leq \theta_0 \text{ v.s. } H_1 : \theta > \theta_0.$$

  Example 10 and the above argument has shown that for any $\theta_0' \leq \theta_0$, the UMP test for the simple v.s. composite hypotheses

$$H_0 : \theta = \theta_0' \text{ v.s. } H_1 : \theta > \theta_0$$

  rejects H_0 if

$$\bar{X}_n \leq c_\alpha(\theta_0') = \frac{Z_\alpha}{\theta_0'\sqrt{n}} + \frac{1}{\theta_0'},$$

  Since we are okay on the power side, now what is left to show is that such rejection region can control the size. Consider the rejection region for $\theta_0' = \theta_0$:

$$\sup_{\theta_0' \leq \theta_0} P_{\theta_)'}(\bar{X}_n \leq \frac{Z_\alpha}{\theta_0\sqrt{n}} + \frac{1}{\theta_0}) \leq \sup_{\theta_0' \leq \theta_0} P_{\theta_)'}(\bar{X}_n \leq \frac{Z_\alpha}{\theta_)'\sqrt{n}} + \frac{1}{\theta_)'}) = \alpha.$$

  This completes the argument.

3 Asymptotic Properties of LR, Wald and LM Tests

Asymptotic Properties of LR Test

- In previous sections we learned the likelihood ratior (LR) test and that it is the UMP test in some important cases. There, we find the critical region (critical value) of the test by exploiting the *finite sample properties* of each particular population.

- LR test can be widely applied, even to the cases where the finite sample distribution of the corresponding test statistic is hard to derive. Now we study the asymptotic distribution of the LR test.

- Let $X_1, \ldots, X_n$ be a random sample from a population $f(x|\theta)$. Consider the general hypotheses:

$$H_0 : h(\theta) = 0 \text{ versus } H_1 : h(\theta) \neq 0,$$

where $h(\cdot)$ is a known r-vector valued function, i.e. $h(\theta) \in \mathbb{R}^r$. If θ is a $k \times 1$ vector, we assume $r \leq k$.

- **Theorem 2** *Suppose some regularity conditions for the MLE hold. Then under the null hypothesis,*

$$LR_n \equiv -2\log\lambda(\mathbf{X}) \equiv -2\log\left[\frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{X})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{X})}\right] \xrightarrow{d.} \chi_r^2 \text{ as } n \to \infty,$$

where r is the number of restrictions imposed on the parameter by the null hypothesis.

- Using the result of the theorem, an approximate level α LR test is to reject H_0 if

$$LR_n \equiv -2\log\lambda(\mathbf{X}) > \chi^2_{r,1-\alpha}.$$

- This theorem gives the same rejection region as in Example 6, but it does not require the population to be normal. It is also similar to Example 7, in the sense that χ_1^2 is the limit of the $F_{1,n-1}$ distribution derived there as $n \to \infty$. Even though Examples 6 and 7 both have normal population, Theorem 2 can be applied to other distributions as well.

- *Proof of Theorem 2.* Here we sketch the proof only for a special case:

$$H_0 : \theta = \theta_0 \text{ v.s. } H_1 : \theta \neq \theta_0.$$

Note that $k = r$ in this case. By the definition of the likelihood function, we have

$$
\begin{aligned}
-2\log\lambda(\mathbf{X}) &= -2\log\left[\frac{\sup_{\theta\in\Theta_0} L(\theta|\mathbf{X})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{X})}\right] \\
&= -2\log\left(\frac{\Pi_{i=1}^n f(X_i|\theta_0)}{\Pi_{i=1}^n f(X_i|\hat{\theta}_n)}\right) \\
&= -2\left(\sum_{i=1}^n \log f(X_i|\theta_0) - \sum_{i=1}^n \log f(X_i|\hat{\theta}_n)\right),
\end{aligned}
$$

where $\hat{\theta}_n$ denotes the MLE of θ. $\hat{\theta}_n$ satisfies the following FOC:

$$
\sum_{i=1}^n \frac{\partial \log f(X_i|\hat{\theta}_n)}{\partial\theta} = 0.
$$

Applying Taylor expansion to $\sum_{i=1}^n \log f(X_i|\theta_0)$ around $\hat{\theta}_n$ gives

$$
\begin{aligned}
\sum_{i=1}^n \log f(X_i|\theta_0) =& \sum_{i=1}^n \log f(X_i|\hat{\theta}_n) + \underbrace{\sum_{i=1}^n \frac{\partial \log f(X_i|\hat{\theta}_n)}{\partial\theta}}_{=0}(\theta_0 - \hat{\theta}_n) \\
&+ \frac{1}{2}(\theta_0 - \hat{\theta}_n)' \sum_{i=1}^n \frac{\partial^2 \log f(X_i|\tilde{\theta}_n)}{\partial\theta\partial\theta'}(\theta_0 - \hat{\theta}_n),
\end{aligned}
$$

where $\tilde{\theta}_n$ is some value between θ_0 and $\hat{\theta}_n$. Plugging this into the likelihood ratio, we get

$$
-2\log\lambda(\mathbf{X}) = \sqrt{n}(\theta_0 - \hat{\theta}_n)' \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log f(X_i|\tilde{\theta}_n)}{\partial\theta\partial\theta'}\right] \sqrt{n}(\theta_0 - \hat{\theta}_n).
$$

Suppose the following results hold under the null (without proof):

1. the MLE has the asymptotic distribution

$$
\sqrt{n}(\theta_0 - \hat{\theta}_n) \xrightarrow{d.} \mathcal{N}(0, I^{-1}(\theta_0)),
$$

where $I(\theta_0) \equiv -\mathbb{E}(\frac{\partial^2 \log f(X_i|\theta_0)}{\partial\theta\partial\theta'})$ is the information matrix;

2. by the uniform law of large numbers and the consistency of $\hat{\theta}_n$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(X_i|\tilde{\theta}_n)}{\partial\theta\partial\theta'} \xrightarrow{p.} \mathbb{E}\left[\frac{\partial^2 \log f(X|\tilde{\theta}_n)}{\partial\theta\partial\theta'}\right] = -I(\theta_0).$$

Combining these results and by the continuous mapping theorem and the Slutszky's theorem, we get

$$LR_n \equiv -2\log\lambda(\mathbf{X}) \xrightarrow{d.} S'I(\theta_0)S \stackrel{d.}{=} Z'Z \stackrel{d.}{=} \chi^2_r,$$

where $S \sim \mathcal{N}(0, I^{-1}(\theta_0))$ and $Z \equiv I^{1/2}(\theta_0)S \sim \mathcal{N}(0, I_r)$.

- Note that the LR test requires two estimators: the unrestricted estimator $\hat{\theta}_n$ and the retricted estimator θ_0 (the one under H_0).

Asymptotic Properties of Wald Test

- Based on the asymptotic properties of the estimators, two other tests are also frequently used. We will learn them in this and the next subsections.

- Again, consider the general hypotheses:

$$H_0 : h(\theta_0) = 0 \text{ versus } H_1 : h(\theta_0) \neq 0,$$

where $h(\cdot)$ is a known function that takes values in $\mathbb{R}^r$ and is continuously differentiable in a neighborhood of the true parameter value θ_0. **Wald test** is based on the idea that if the asymptotic distribution of some estimator $\hat{\theta}_n$ (not necessarily the MLE) is known, then the delta method gives us the asymptotic distribution of $h(\hat{\theta}_n)$.

- Formally, suppose under the null hypothesis an estimator $\hat{\theta}_n$ has the asymptotic distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d.} \mathcal{N}(0, V(\theta_0)),$$

for some $k \times k$ asymptotic covariance matrix $V(\theta_0)$. Then by the delta method,

$$\sqrt{n}[h(\hat{\theta}_n) - h(\theta_0)] \xrightarrow{d.} \mathcal{N}(0, H(\theta_0)V(\theta_0)H'(\theta_0)),$$

where $H(\theta_0) \equiv \partial h(\theta_0)/\partial\theta \neq 0$. Suppose a consistent estimator $\hat{V}_n(\hat{\theta}_n)$ of $V(\theta_0)$ is available, then by the continuous mapping theorem, we have

$$H(\hat{\theta}_n)\hat{V}_n(\hat{\theta}_n)H'(\hat{\theta}_n) \xrightarrow{p.} H(\theta_0)V(\theta_0)H'(\theta_0).$$

Then by the continuous mapping theorem, we have that the Wald test statistic

$$W_n \equiv \sqrt{n}h'(\hat{\theta}_n)[H(\hat{\theta}_n)\hat{V}_n(\hat{\theta}_n)H'(\hat{\theta}_n)]^{-1}\sqrt{n}h(\hat{\theta}_n) \xrightarrow{d.} \chi_r^2.$$

Therefore, a level α Wald test is to reject H_0 if

$$W_n > \chi^2_{r,1-\alpha}.$$

- Note that the Wald test requires only estimator: the unrestricted estimator $\hat{\theta}_n$.

Asymptotic Properties of Lagrange Multiplier (LM) Test

- Suppose we want to test the hypotheses

$$H_0 : \theta = \theta^* \text{ v.s. } H_1 : \theta \neq \theta^*.$$

Consider the restricted estimator (under H_0)

$$\tilde{\theta}_n \equiv \arg\max_{\theta} \sum_{i=1}^{n} m(X_i, \theta) + \lambda_n \cdot (\theta - \theta^*),$$

where m could be the likelihood function, or a negative loss function, and λ_n is the Lagrange multiplier of this constrained optimization problem. Assume that we have an interior solution, then note that the following FOC hold (with respect to θ and λ_n respectively):

$$\sum_{i=1}^{n} \frac{\partial m(X_i, \tilde{\theta}_n)}{\partial\theta} = \lambda_n,$$
$$\tilde{\theta}_n = \theta^*,$$

which imply

$$\lambda_n = \sum_{i=1}^{n} \frac{\partial m(X_i, \theta^*)}{\partial\theta} \tag{4}$$

Under H_0, $\theta = \theta^*$, i.e. the restriction is not binding, so the Lagrange multiplier should be zero. This means that $\mathbb{E}(\lambda_n) = \mathbb{E}\left[\sum_{i=1}^n \frac{\partial m(X_i,\theta^*)}{\partial \theta}\right] = 0$. Even taking into account of the sampling errors, λ_n should be close to zero if H_0 is true. This is the intuition of the **Lagrange Multiplier (LM) Test**. Under H_0 and by the CLT, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\partial m(X_i,\theta^*)}{\partial \theta} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\frac{\partial m(X_i,\theta^*)}{\partial \theta} - \mathbb{E}\left(\frac{\partial m(X_i,\theta^*)}{\partial \theta}\right)\right] \\ &\xrightarrow{d.} \mathcal{N}(0, V(\theta^*)),\end{aligned}$$

for some asymptotic covariance matrix $V(\theta^*)$. Suppose a consistent estimator $\hat{V}_n(\tilde{\theta}_n)$ of $V(\theta^*)$ is available ($\tilde{\theta}_n$ is consistent under H_0). Then by the continuous mapping theorem and equation (4), we have that the LM test statistic

$$\begin{aligned}LM_n &\equiv \frac{1}{n}\lambda_n'[\hat{V}_n(\hat{\theta}_n)]^{-1}\lambda_n \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\partial m(X_i,\theta^*)}{\partial \theta'}[\hat{V}_n(\tilde{\theta}_n)]^{-1}\sum_{i=1}^n \frac{\partial m(X_i,\theta^*)}{\partial \theta} \\ &\xrightarrow{d.} \chi_r^2.\end{aligned}$$

Therefore, a level α Lagrange multiplier test is to reject H_0 if

$$LM_n > \chi^2_{r,1-\alpha}.$$

- Note that the LM test requires only estimator: the retricted estimator θ_0 (the one under H_0).
- Figure 4.8 in Greene's Econometric Analysis (copied here) illustrates the idea of these three tests.
- The three asymptotic tests introduced in this section can be widely used. We will discuss them again when talking about regression models.

4 Exercises

1. Prove equation (1).

2. Let $X_1, \ldots, X_n$ be a random sample from a population $\mathcal{N}(\mu, \sigma^2)$ where μ is unknown. Suppose we want to test

$$H_0 : \mu \leq \mu_0 \text{ v.s. } H_1 : \mu > \mu_0.$$

   (a) If σ^2 is known, find the likelihood ratio test statistic;

   (b) Show that the LR test based on the statistic found in (a) is equivalent to the test that rejects H_0 when

$$\bar{X} > \mu_0 + z_{1-\alpha}\sqrt{\sigma^2/n},$$

   at the significance level α;

   (c) Provide the argument for that the test in (b) is a UMP test;

   (d) If σ^2 is unknown, find the LR test statistic;

   (e) Show that the LR test based on the statistic found in (d) is equivalent to the test that rejects H_0 when

$$\bar{X} > \mu_0 + t_{n-1,1-\alpha}\sqrt{S^2/n}.$$

3. Let X be a random variable whose pmf under $H_0 : \theta = \theta_0$ and under $H_1 : \theta = \theta_1$ are given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

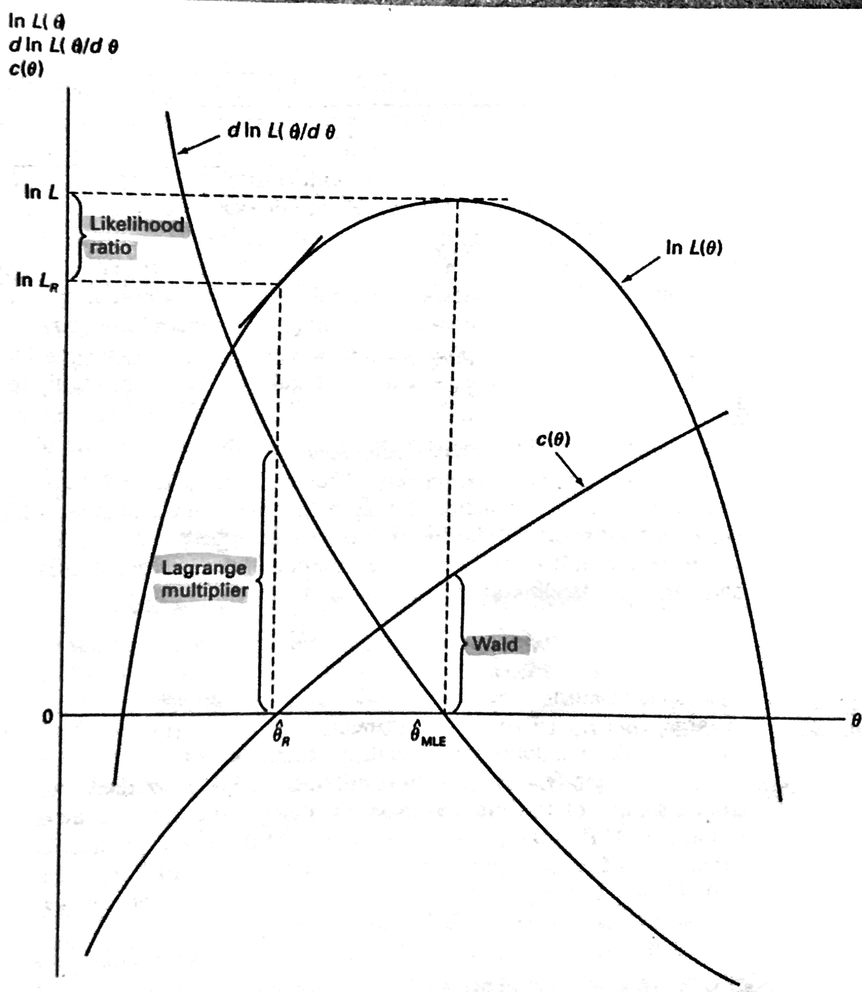
   Use the Neyman-Pearson lemma to find the UMP test for H_0 v.s. H_1 with size $\alpha = 0.04$. Compute the probability of Type II error for the test that you found.

4. Show (3) using the Neyman-Pearson lemma. Find the power function for the UMP test in the example.

5. Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(0, \sigma^2)$ population. Show that there is not a UMP test for the hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 \neq \sigma_0^2.$$

FIGURE 4.8 Three Bases for Hypothesis Tests



[21]See Buse (1982). Note that the scale of the vertical axis would be different for each curve. As such, the points of intersection have no significance.