Lecture Notes 10
Hypothesis Testing

1 Introduction

(See Chapter 10).

Null hypothesis: \( H_0 : \theta \in \Theta_0 \)
Alternative hypothesis: \( H_1 : \theta \in \Theta_1 \)
where \( \Theta_0 \cap \Theta_1 = \emptyset \).

Example 1 \( X_1, \ldots, X_n \sim \text{Bernoulli}(p) \).

\[
H_0 : p = \frac{1}{2} \quad H_1 : p \neq \frac{1}{2}.
\]

The question is not whether \( H_0 \) is true or false. The question is whether there is sufficient evidence to reject \( H_0 \), much like a court case.

Our possible actions are: reject \( H_0 \) or retain (don’t reject) \( H_0 \).

<table>
<thead>
<tr>
<th>Decision</th>
<th>Retain ( H_0 )</th>
<th>Reject ( H_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) true</td>
<td>( \sqrt{} )</td>
<td>Type I error (false positive)</td>
</tr>
<tr>
<td>( H_1 ) true</td>
<td>Type II error (false negative)</td>
<td>( \sqrt{} )</td>
</tr>
</tbody>
</table>

Warning: Hypothesis testing should only be used when it is appropriate. Often times, people use hypothesis testing when it would be much more appropriate to use confidence intervals (which is the next topic).

2 Constructing Tests

1. Choose a test statistic \( W = W(X_1, \ldots, X_n) \).
2. Choose a rejection region \( R \).
3. If \( W \in R \) we reject \( H_0 \) otherwise we retain \( H_0 \).
Example 2 $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$.

$$H_0 : p = \frac{1}{2} \quad H_1 : p \neq \frac{1}{2}.$$  

Let $W = n^{-1} \sum_{i=1}^n X_i$. Let $R = \{ x^n : |w(x^n) - 1/2| > \delta \}$. So we reject $H_0$ if $|W - 1/2| > \delta$.

We need to choose $W$ and $R$ so that the test has good statistical properties. We will consider the following tests:

1. Neyman-Pearson Test
2. Wald test
3. Likelihood Ratio Test (LRT)
4. the permutation test

Before we discuss these methods, we first need to talk about how we evaluate tests.

## 3 Error Rates and Power

Suppose we reject $H_0$ when $X^n = (X_1, \ldots, X_n) \in R$. Define the power function by

$$\beta(\theta) = P_{\theta}(X^n \in R).$$

We want $\beta(\theta)$ to be small when $\theta \in \Theta_0$ and we want $\beta(\theta)$ to be large when $\theta \in \Theta_1$.

The general strategy is:

1. Fix $\alpha \in [0, 1]$.
2. Now try to maximize $\beta(\theta)$ for $\theta \in \Theta_1$ subject to $\beta(\theta) \leq \alpha$ for $\theta \in \Theta_0$.

We need the following definitions. A test is size $\alpha$ if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$  

A test is level $\alpha$ if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$  

A size $\alpha$ test and a level $\alpha$ test are almost the same thing. The distinction is made because sometimes we want a size $\alpha$ test and we cannot construct a test with exact size $\alpha$ but we can construct one with a smaller error rate.
Example 3 \(X_1, \ldots, X_n \sim N(\theta, \sigma^2)\) with \(\sigma^2\) known. Suppose
\[H_0 : \theta = \theta_0, \quad H_1 : \theta > \theta_0.\]
This is called a **one-sided alternative**. Suppose we reject \(H_0\) if \(W > c\) where
\[W = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}.\]
Then
\[
\beta(\theta) = P_\theta \left( \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} > c \right)
= P_\theta \left( \frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
= P \left( Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
= 1 - \Phi \left( c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
\]
where \(\Phi\) is the cdf of a standard Normal. Now
\[
\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = 1 - \Phi(c).
\]
To get a size \(\alpha\) test, set \(1 - \Phi(c) = \alpha\) so that
\[c = z_\alpha\]
where \(z_\alpha = \Phi^{-1}(1 - \alpha)\). Our test is: reject \(H_0\) when
\[W = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} > z_\alpha.\]

Example 4 \(X_1, \ldots, X_n \sim N(\theta, \sigma^2)\) with \(\sigma^2\) known. Suppose
\[H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.\]
This is called a **two-sided** alternative. We will reject \(H_0\) if \(|W| > c\) where \(W\) is defined as before. Now
\[
\beta(\theta) = P_\theta(W < -c) + P_\theta(W > c)
= P_\theta \left( \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} < -c \right) + P_\theta \left( \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} > c \right)
= P \left( Z < -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + P \left( Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
= \Phi \left( -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + 1 - \Phi \left( c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
= \Phi \left( -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + \Phi \left( -c - \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
\]
since \( \Phi(-x) = 1 - \Phi(x) \). The size is
\[
\beta(\theta_0) = 2\Phi(-c).
\]
To get a size \( \alpha \) test we set \( 2\Phi(-c) = \alpha \) so that \( c = -\Phi^{-1}(\alpha/2) = \Phi^{-1}(1 - \alpha/2) = z_{\alpha/2} \). The test is: reject \( H_0 \) when
\[
|W| = \left| \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}.
\]

4 The Neyman-Pearson Test

(Not in the book.) Let \( C_\alpha \) denote all level \( \alpha \) tests. A test in \( C_\alpha \) with power function \( \beta \) is uniformly most powerful (UMP) if the following holds: if \( \beta' \) is the power function of any other test in \( C_\alpha \) then \( \beta(\theta) \leq \beta'(\theta) \) for all \( \theta \in \Theta_1 \).

Consider testing \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta = \theta_1 \). (Simple null and simple alternative.)

**Theorem 5** Suppose we set
\[
R = \left\{ x = (x_1, \ldots, x_n) : \frac{p(X_1, \ldots, X_n; \theta_1)}{p(X_1, \ldots, X_n; \theta_0)} > k \right\} = \left\{ x^n : \frac{L(\theta_1)}{L(\theta_0)} > k \right\}
\]
where \( k \) is chosen so that
\[
P_{\theta_0}(X^n \in R) = \alpha.
\]
In other words, reject \( H_0 \) if
\[
\frac{L(\theta_1)}{L(\theta_0)} > k.
\]
This test is a UMP level \( \alpha \) test.

5 The Wald Test

Let
\[
W = \frac{\hat{\theta}_n - \theta_0}{se}.
\]
Under the usual conditions we have that under \( H_0 \), \( W \sim N(0,1) \). Hence, an asymptotic level \( \alpha \) test is to reject when \( |W| > z_{\alpha/2} \).

For example, with Bernoulli data, to test \( H_0 : p = p_0 \),
\[
W = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-p)}{n}}},
\]
You can also use
\[
W = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},
\]
In other words, to compute the standard error, you can replace \( \theta \) with an estimate \( \hat{\theta} \) or by the null value \( \theta_0 \).
6 The Likelihood Ratio Test (LRT)

This test is simple: reject $H_0$ if $\lambda(x^n) \leq c$ where

$$\lambda(x^n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

where $\hat{\theta}_0$ maximizes $L(\theta)$ subject to $\theta \in \Theta_0$.

Example 6 $X_1, \ldots, X_n \sim N(\theta,1)$. Suppose

$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$

After some algebra,

$$\lambda = \exp\left\{-\frac{n}{2}(X_n - \theta_0)^2\right\}.$$ 

So

$$R = \{x : \lambda \leq c\} = \{x : |X - \theta_0| \geq c\}$$

where $c' = \sqrt{-2\log c/n}$. Choosing $c'$ to make this level $\alpha$ gives: reject if $|W| > z_{\alpha/2}$ where $W = \sqrt{n}(X - \theta_0)$ which is the test we constructed before.

Example 7 $X_1, \ldots, X_n \sim N(\theta,\sigma^2)$. Suppose

$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$

Then

$$\lambda(x^n) = \frac{L(\theta_0, \hat{\sigma}_0)}{L(\hat{\theta}, \hat{\sigma})}$$

where $\hat{\sigma}_0$ maximizes the likelihood subject to $\theta = \theta_0$. In the homework, you will prove that $\lambda(x^n) < c$ corresponds to rejecting when $|T_n| > k$ for some constant $k$ where

$$T_n = \frac{X_n - \theta_0}{S/\sqrt{n}}.$$

Under $H_0$, $T_n$ has a $t$-distribution with $n - 1$ degrees of freedom. So the final test is: reject $H_0$ if

$$|T_n| > t_{n-1,\alpha/2}.$$ 

This is called Student’s $t$-test. It was invented by William Gosset working at Guiness Breweries and writing under the pseudonym Student.

Theorem 8 Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ where $\theta \in \mathbb{R}$. Under $H_0$,

$$-2\log \lambda(X^n) \sim \chi_1^2.$$ 

Hence, if we let $W_n = -2\log \lambda(X^n)$ then

$$P_{\theta_0}(W > \chi_{1,\alpha}^2) \to \alpha$$

as $n \to \infty$. 

5
**Proof.** Using a Taylor expansion:

\[
\ell(\theta) \approx \ell(\hat{\theta}) + \ell'(\hat{\theta})(\theta - \hat{\theta}) + \ell''(\hat{\theta})(\theta - \hat{\theta})^2 = \ell(\hat{\theta}) + \ell''(\hat{\theta})(\theta - \hat{\theta})^2
\]

and so

\[
-2 \log \lambda(x^n) = 2\ell(\hat{\theta}) - 2\ell(\theta_0) \\
\approx 2\ell(\hat{\theta}) - 2\ell(\hat{\theta}) - \ell''(\hat{\theta})(\theta - \hat{\theta})^2 = -\ell''(\hat{\theta})(\theta - \hat{\theta})^2 \\
= -\frac{\ell''(\hat{\theta})}{I_n(\theta_0)} I_n(\hat{\theta}_0)(\sqrt{n}(\hat{\theta} - \theta_0))^2 = A_n \times B_n.
\]

Now \(A_n \overset{P}{\rightarrow} 1\) by the WLLN and \(\sqrt{B_n} \overset{d}{\rightarrow} N(0, 1)\). The result follows by Slutsky’s theorem.

□

**Example 9** \(X_1, \ldots, X_n \sim \text{Poisson}(\lambda)\). We want to test \(H_0 : \lambda = \lambda_0\) versus \(H_1 : \lambda \neq \lambda_0\).

Then

\[
-2 \log \lambda(x^n) = 2n[(\lambda_0 - \hat{\lambda}) - \hat{\lambda} \log(\lambda_0/\hat{\lambda})].
\]

We reject \(H_0\) when \(-2 \log \lambda(x^n) > \chi^2_{1,\alpha}\).

Now suppose that \(\theta = (\theta_1, \ldots, \theta_k)\). Suppose that \(H_0\) fixes some of the parameters. Then

\[
-2 \log \lambda(X^n) \sim \chi^2_{\nu}
\]

where

\[
\nu = \dim(\Theta) - \dim(\Theta_0).
\]

**Example 10** Consider a multinomial with \(\theta = (p_1, \ldots, p_5)\). So

\[
L(\theta) = p_1^{y_1} \cdots p_5^{y_5}.
\]

Suppose we want to test

\(H_0 : p_1 = p_2 = p_3\) and \(p_4 = p_5\)

versus the alternative that \(H_0\) is false. In this case

\[
\nu = 4 - 1 = 3.
\]

The LRT test statistic is

\[
\lambda(x^n) = \frac{\prod_{i=1}^{5} p_0^{Y_i}}{\prod_{i=1}^{5} \hat{p}_j^{Y_i}}
\]

where \(\hat{p}_j = Y_j/n, \hat{p}_{10} = \hat{p}_{20} = \hat{p}_{30} = (Y_1 + Y_2 + Y_3)/n, \hat{p}_{40} = \hat{p}_{50} = (1 - 3\hat{p}_{10})/2\). These calculations are on p 491. Make sure you understand them. Now we reject \(H_0\) if \(-2\lambda(X^n) > \chi^2_{3,\alpha}\). □
7 p-values

When we test at a given level \( \alpha \) we will reject or not reject. It is useful to summarize what levels we would reject at and what levels we would not reject at.

The p-value is the smallest \( \alpha \) at which we would reject \( H_0 \).

In other words, we reject at all \( \alpha \geq p \). So, if the p-value is 0.03, then we would reject at \( \alpha = 0.05 \) but not at \( \alpha = 0.01 \).

Hence, to test at level \( \alpha \) when \( p < \alpha \).

**Theorem 11** Suppose we have a test of the form: reject when \( W(X^n) > c \). Then the p-value when \( X^n = x^n \) is

\[
p(x^n) = \sup_{\theta \in \Theta_0} P_\theta(W(X^n) \geq W(x^n)).
\]

**Example 12** \( X_1, \ldots, X_n \sim N(\theta, 1) \). Test that \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \). We reject when \( |W| \) is large, where \( W = \sqrt{n}(\bar{X}_n - \theta_0) \). So

\[
p = P_{\theta_0}(|\sqrt{n}(\bar{X}_n - \theta_0)| > w) = P(|Z| > w) = 2\Phi(-|w|).
\]

**Theorem 13** Under \( H_0 \), \( p \sim \text{Unif}(0, 1) \).

**Important.** Note that \( p \) is NOT equal to \( P(H_0|X_1, \ldots, X_n) \). The latter is a Bayesian quantity which we will discuss later.

8 The Permutation Test

This is a very cool test. It is distribution free and it does not involve any asymptotic approximations.

Suppose we have data

\[
X_1, \ldots, X_n \sim F
\]

and

\[
Y_1, \ldots, Y_m \sim G.
\]

We want to test:

\[ H_0 : F = G \quad \text{versus} \quad H_1 : F \neq G. \]

Let

\[ Z = (X_1, \ldots, X_n, Y_1, \ldots, Y_m). \]

Create labels

\[ L = (1, \ldots, 1, 2, \ldots, 2). \]

A test statistic can be written as a function of \( Z \) and \( L \). For example, if

\[ W = |\bar{X}_n - \bar{Y}_n| \]
then we can write
\[
W = \left| \frac{\sum_{i=1}^{N} Z_i I(L_i = 1)}{\sum_{i=1}^{N} I(L_i = 1)} - \frac{\sum_{i=1}^{N} Z_i I(L_i = 2)}{\sum_{i=1}^{N} I(L_i = 2)} \right|
\]
where \( N = n + m \). So we write \( W = g(L, Z) \).

Define
\[
p = \frac{1}{N!} \sum_{\pi} I(g(L_{\pi}, Z) > g(L, Z))
\]
where \( L_{\pi} \) is a permutation of the labels and the sum is over all permutations. Under \( H_0 \), permuting the labels does not change the distribution. In other words, \( g(L, Z) \) has an equal chance of having any rank among all the permuted values. That is, under \( H_0 \), \( \approx \text{Unif}(0, 1) \) and if we reject when \( p < \alpha \), then we have a level \( \alpha \) test.

Summing over all permutations is infeasible. But it suffices to use a random sample of permutations. So we do this:

1. Compute a random permutation of the labels and compute \( W \). Do this \( K \) times giving values \( W_1, \ldots, W_K \).

2. Compute the p-value
\[
\frac{1}{K} \sum_{j=1}^{K} I(W_j > W).
\]