Lecture Notes 9
Hypothesis Testing

1 Introduction

Definition 1 A hypothesis is a statement about a population parameter.

Null hypothesis: $H_0 : \theta \in \Theta_0$

Alternative hypothesis: $H_1 : \theta \in \Theta_1$

where $\Theta_0 \cap \Theta_1 = \emptyset$.

Example 2 $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$.

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

• Population: students at U of M who are taking Stats 511.

• Population parameter $p$ describes the probability of a student who is taking Stats 511 saying “No” regarding his/her preference of seeing proof.

• If I poll students regarding whether they like to see proofs or not:
  – Null hypothesis states that the probability each of them says “No” is $1/2$.
  – Alternative hypothesis states that they tend to say “No” more often than “Yes”.

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) $H_0$.

\[
\begin{array}{c|c|c}
 & \text{Decision} & \\
\hline
 & \text{Retain } H_0 & \text{Reject } H_0 \\
\hline
H_0 \text{ true} & \checkmark & \text{Type I error (false positive)} \\
H_1 \text{ true} & \text{Type II error (false negative)} & \checkmark \\
\end{array}
\]
2 Constructing Tests

Usually we do the following:

1. Choose a test statistic $W(X^n) = W(X_1, \ldots, X_n)$.
2. Choose a rejection region $R$.
3. If $X^n \in R$ we reject $H_0$ otherwise we retain $H_0$.

Usually, the rejection region $R$ is of the form

$$R = \{ x^n : W(x^n) > c \}$$

where $c$ is a critical value. Clearly, we need to choose $W$ and $R$ so that the test has good statistical properties. Suppose we reject $H_0$ when $X^n = (X_1, \ldots, X_n) \in R$.

**Definition 3** (CB Definition 8.3.1) The power function of a hypothesis test with rejection region $R$ is the function of $\theta$ defined by

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

**Interpretation:** $\beta$ represents the “power” of making the correct decision by rejecting the null hypothesis $H_0$ when it is wrong: that is,

for $\theta \in \Theta_1$, $\beta(\theta) = P_{\theta}(X^n \in R) = 1 - \text{Type II error}$.

Based on the power function, we define

1. Type I Error: $\beta(\theta) = P_{\theta}(X^n \in R)$ if $\theta \in \Theta_0$
2. Type II Error: $1 - \beta(\theta) = P_{\theta}(X^n \in R^c)$ if $\theta \in \Theta_1$ where $R^c$ is the complement of $R$.

Put in words,

1. Type I Error is the probability of rejecting $H_0$ when $H_0$ is true.
2. Type II Error is the probability of retaining (accepting) $H_0$ when $H_1$ is true.

The ideal power function is 0 for all $\theta \in \Theta_0$ and is 1 for all $\theta \in \Theta_1$ so that errors of both types would be 0. In reality, we want $\beta(\theta)$ to be small when $\theta \in \Theta_0$ and we want $\beta(\theta)$ to be large when $\theta \in \Theta_1$. 
Example 4 (CB Example 8.3.2) \(X_1, \ldots, X_n \sim \text{Bernoulli}(p)\) Test 1.

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

In the first test we construct in class, we let \(n = 5\) and define

\[
W(X^n) = \sum_{i=1}^{n} X_i \quad \text{and} \quad R_1 = \{x^n : w(x^n) = \sum_{i=1}^{n} x_i = 5\}
\]

Following Example 2 above, we reject \(H_0\) if all 5 student say “No”. Based on the polling results in my class in the morning session being “No, No, No, Yes, Yes” we decide to retain \(H_0\) based on Test 1.

\[
\beta_1(\theta) = P_\theta(X^n \in R) = P_\theta\left(\sum_{i=1}^{5} X_i = 5\right) = \theta^5.
\]

For this test we see a small Type I error, but a large Type II error for most of \(\theta \in H_1\).

Example 5 \(X_1, \ldots, X_n \sim \text{Bernoulli}(p)\) Test 2. Following Example 4 and consider

\[
W(X^n) = \sum_{i=1}^{n} X_i \quad \text{and} \quad R_2 = \{x^n : w(x^n) = \sum_{i=1}^{n} x_i = 3, 4, \text{ or } 5\}
\]

Following Example 2 above, we reject \(H_0\) given that three students have said “No”.

\[
\beta_2(\theta) = P_\theta(X^n \in R) = P_\theta\left(\sum_{i=1}^{5} X_i = 3, 4, \text{ or } 5\right)
= \binom{5}{3} \theta^3 (1 - \theta)^2 + \binom{5}{4} \theta^4 (1 - \theta) + \binom{5}{5} \theta^5
\]

For this second test we achieve a smaller Type II error for every \(\theta \in \Theta_0\), but a larger Type I error for every \(\theta \in \Theta_0\). If a choice is to be made between \(R_1\) and \(R_2\), the experimenter must decide which error structure, that described by \(\beta_1(\theta)\) or that described by \(\beta_2(\theta)\), is more acceptable.

Exercise 1. Plot the power functions for \(\beta_1\) and \(\beta_2\).

We will consider the following tests:

1. Likelihood Ratio Test (LRT)
2. Neyman-Pearson Test
3. Wald Test

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

Definition 6  For 0 ≤ α ≤ 1, a test with power function β(θ) is a size α test if

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

Definition 7  For 0 ≤ α ≤ 1, a test with power function β(θ) is a level α test if

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

A size α test and a level α test are almost the same thing. The distinction is made because sometimes we want a size α test and we cannot construct a test with exact size α but we can construct one with a smaller error rate, that is a test that is of level α. Hence the set of level α tests contains the set of size α tests.

The general strategy is: fix \( \alpha \in [0,1] \). Now try to

\[
\text{maximize } \beta(\theta) \text{ for } \theta \in \Theta_1 \\
\text{subject to } \beta(\theta) \leq \alpha \text{ for } \theta \in \Theta_0
\]

with typical choices being \( \alpha = .01, .05, .10 \). That is, the experimenter controls the Type I error. If this approach is taken, then the experimenter should specify the null and alternative hypotheses so that it is most important to control the Type I Error probability while maximizing the power of rejecting \( H_0 \) when \( H_1 \) is true.

4 The Likelihood Ratio Test (LRT)

Null hypothesis: \( H_0 : \theta \in \Theta_0 \)

Alternative hypothesis: \( H_1 : \theta \in \Theta_1 \)

where \( \Theta_0 \cap \Theta_1 = \emptyset \) and \( \Theta_0 \cup \Theta_1 = \Theta \).

Definition 8  The Likelihood Ratio Test statistic is:

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

where \( \hat{\theta}_0 \) maximizes \( L(\theta) \) subject to \( \theta \in \Theta_0 \) and \( \hat{\theta} \) is the MLE for \( \theta \) given a sample point \( x^n \).

The Likelihood Ratio Test (LRT) is any test that has a rejection region of the form

\[ R = \{ x^n : \lambda(x^n) \leq c \} \]

where \( c \) is any number satisfying 0 ≤ c ≤ 1.
Example 9 (CB 8.2.2), $X_1, \ldots, X_n \sim N(\theta, 1)$. Suppose

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

After some algebra

$$\lambda(x^n) = \exp \left\{ -\frac{n}{2} (\bar{X}_n - \theta_0)^2 \right\}.$$ 

So

$$R = \{ x^n : \lambda \leq c \} = \{ x^n : |\bar{X}_n - \theta_0| \geq \sqrt{-2 \log c/n} \}.$$ 

As $c$ ranges between 0 and 1, $\sqrt{-2 \log c/n} =: c'$ ranges between 0 and $\infty$. Thus the LRTs are just those tests that reject $H_0 : \theta = \theta_0$ if the sample mean differs from the hypothesized value $\theta_0$ by more than a specified amount.

**Exercise 2.** Choose $c$ (and its corresponding $c'$) to make this LRT a level $\alpha$ test.

**Exercise 3.** Plot the power function for the Normal LRT.

### 5 The LRT on one-sided alternative

Next we consider the following *One-sided Alternative Hypotheses* problem:

Null hypothesis: $H_0 : \theta \leq \theta_0$

Alternative hypothesis: $H_1 : \theta > \theta_0$

Example 10 $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with $\sigma^2$ known. Suppose

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta \in \Theta_1 = \{ \theta : \theta > \theta_0 \}.$$ 

This is called a one-sided alternative.

Suppose the Likelihood Ratio Test (LRT) is defined as: Rejection region

$$R = \{ x^n : \lambda(x^n) \leq c_0 \}$$

where $0 \leq c_0 \leq 1$. This is the same test as (WHY?)

Reject $H_0$ if \[ W > \sqrt{-2 \ln c_0} := c \]

where \[ W = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}. \]
Then

\[
\beta(\theta) = P_\theta \left( \frac{X_n - \theta_0}{\sigma/\sqrt{n}} > c \right)
\]

\[
= P_\theta \left( \frac{X_n - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
\]

\[
= P \left( Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
\]

\[
= 1 - \Phi \left( c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)
\]

where \( \Phi \) is the cdf of a standard Normal. This is plotted on page 384 of Casella & Berger.

Now

\[
\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = 1 - \Phi(c) \quad (\text{why?})
\]

Answer: because \( \beta \) is a monotonically increasing function of \( \theta \); hence \( \theta_0 \) achieves the supremum in above expression.

To get a size \( \alpha \) test, set \( 1 - \Phi(c) = \alpha \) so that

\[
c = z_\alpha, \quad \text{for example: } c = 1.28 \text{ for } \alpha = 0.1,
\]

where \( z_\alpha = \Phi^{-1}(1 - \alpha) \). Our size \( \alpha \) test is: reject \( H_0 \) when

\[
W = \frac{X_n - \theta_0}{\sigma/\sqrt{n}} > z_\alpha.
\]

Note that this test is independent of \( \theta_1 \in \Theta_1 \); that is, the same test maximizes the power for all possible \( \theta_1 > \theta_0 \). Indeed, this test is called the Uniformly Most Powerful test at level \( \alpha \) for the one-sided alternative problem. This is the topic of next lectures.

Exercises II:

1. Convince yourself as \( n \) increases, \( \beta(\theta_1) \) increases for all \( \theta_1 \in \Theta_1 \).

2. Compare the power plot for \( \beta(\theta) \) with your homework problem.

3. Fill up all steps before \( \beta(\theta) \) function as written above.
The Two-sided Alternative Hypotheses problem: revisited

Consider the following Two-sided Alternative Hypotheses problem:

Null hypothesis: $H_0 : \theta = \theta_0$
Alternative hypothesis: $H_1 : \theta \neq \theta_0$

**Example 11** $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ with $\sigma^2$ known. Recall by LRT, we will reject $H_0$ if $|W| > c$ where $W = \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}}$ is defined as before. The size $\alpha$ test is: reject $H_0$ when

$$|W| = \left| \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}.$$

Recall for $Z \sim N(0, 1)$, (Plot the pdf of standard normal distribution here)

$$\beta(\theta) = P_\theta(W < -c) + P_\theta(W > c)$$
$$= P_\theta \left( \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} < -c \right) + P_\theta \left( \frac{\bar{X}_n - \theta_0}{\sigma/\sqrt{n}} > c \right)$$
$$= P \left( Z < -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + P \left( Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)$$
$$= \Phi \left( -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + 1 - \Phi \left( c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)$$
$$= \Phi \left( -c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + \Phi \left( -c - \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right)$$

since $\Phi(-x) = 1 - \Phi(x)$, where $\Phi$ is the cdf of a standard Normal. The size is

$$\beta(\theta_0) = 2\Phi(-c).$$

To get a size $\alpha$ test we set $2\Phi(-c) = \alpha$ so that $c = -\Phi^{-1}(\alpha/2) = \Phi^{-1}(1 - \alpha/2) = z_{\alpha/2}$. 
