Biostat 602 Winter 2017

Lecture Set 16

Hypothesis Testing

Reading: CB Chapter 8

Hypothesis Testing

A *hypothesis* is a statement about a population parameter

Two complementary statements about θ:

- Null hypothesis : $H_0 : \theta \in \Omega_0$
- Alternative hypothesis : $H_1 : \theta \in \Omega_0^c$

$\theta \in \Omega = \Omega \cup \Omega^c$.

Simple and composite hypothesis

Simple hypothesis

Both H_0 and H_1 consist of only one parameter value.

- $H_0 : \theta = \theta_0 \in \Omega_0$
- $H_1 : \theta = \theta_1 \in \Omega_0^c$

Composite hypothesis

One or both of H_0 and H_1 consist more than one parameter values.

- One-sided hypothesis: $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.
- One-sided hypothesis: $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$.
- Two-sided hypothesis: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

An Example of Hypothesis

$$X_1, \cdots, X_n \quad iid \quad \mathcal{N}(\theta, 1)$$

Let X_i denote the change in blood pressure after a treatment.

$$\begin{aligned} H_0 &: \theta = 0 \qquad \text{(no effect)} \\ H_1 &: \theta \neq 0 \qquad \text{(some effect)} \end{aligned}$$

Two-sided composite hypothesis.

Another Example of Hypothesis

- Let θ denote the proportion of defective items from a machine.
- One may want the proportion to be less than a specified maximum acceptable proportion θ_0.
- We want to test whether the products produced by the machine is acceptable.

$$H_0 : \theta \leq \theta_0 \qquad \text{(acceptable)}$$

$$H_1 : \theta > \theta_0 \qquad \text{(unacceptable)}$$

Hypothesis Testing Procedure

A hypothesis testing procedure is a rule that specifies:

1. For which sample points H_0 is accepted as true (the subset of the sample space for which H_0 is accepted is called the acceptable region).
2. For which sample points H_0 is rejected and H_1 is accepted as true (the subset of sample space for which H_0 is rejected is called the rejection region or critical region).

Rejection region (R) on a hypothesis is usually defined through a test statistic $W(\mathbf{X})$. For example,

$$
\begin{aligned}
R_1 &= \{\mathbf{x} : W(\mathbf{x}) > c, \mathbf{x} \in \mathcal{X}\} \\
R_2 &= \{\mathbf{x} : W(\mathbf{x}) \leq c, \mathbf{x} \in \mathcal{X}\}
\end{aligned}
$$

Example of hypothesis testing

X_1, X_2, X_3 i.i.d. *Bernoulli*(p). Consider hypothesis tests

$$
\begin{aligned}
H_0 &: p \leq 0.5 \\
H_1 &: p > 0.5
\end{aligned}
$$

- Test 1 : Reject H_0 if $\mathbf{x} \in \{(1,1,1)\}$
  $\iff$ rejection region $= \{(1,1,1)\}$
  $\iff$ rejection region $= \{\mathbf{x} : \sum x_i > 2\}$
- Test 2 : Reject H_0 if $\mathbf{x} \in \{(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}$
  $\iff$ rejection region $= \{(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}$
  $\iff$ rejection region $= \{\mathbf{x} : \sum x_i > 1\}$

Example Let $X_1, \cdots, X_n$ be change in blood pressure after a treatment.

$$\begin{aligned} H_0 &: \theta = 0 \\ H_1 &: \theta \neq 0 \end{aligned}$$

An example rejection region $R = \left\{ \mathbf{x} : \frac{\overline{x}}{s_{\mathbf{x}}/\sqrt{n}} > 3 \right\}$.

Decision

Truth		Accept H_0	Reject H_0
	H_0	Correct Decision	Type I error
	H_1	Type II error	Correct Decision

Type I and Type II error

Type I error

If $\theta \in \Omega_0$ (if the null hypothesis is true), the probability of making a type I error is

$$\Pr(\mathbf{X} \in R | \theta)$$

Type II error

If $\theta \in \Omega_0^c$ (if the alternative hypothesis is true), the probability of making a type II error is

$$\Pr(\mathbf{X} \notin R | \theta) = 1 - \Pr(\mathbf{X} \in R | \theta)$$

Power function

Definition: The power function of a hypothesis test with rejection region R is the function of θ defined by

$$\beta(\theta) = \Pr(\mathbf{X} \in R | \theta) = \Pr(\text{reject } H_0 | \theta)$$

If $\theta \in \Omega_0^c$ (alternative is true), the probability of rejecting H_0 is called the power of test for this particular value of θ.

- Probability of type I error $= \beta(\theta)$ if $\theta \in \Omega_0$.
- Probability of type II error $= 1 - \beta(\theta)$ if $\theta \in \Omega_0^c$.

An ideal test should have power function satisfying $\beta(\theta) = 0$ for all $\theta \in \Omega_0$, $\beta(\theta) = 1$ for all $\theta \in \Omega_0^c$, which is typically not possible in practice.

Example 1: Let $X_1, X_2, \cdots, X_n$ i.i.d. $Bernoulli(\theta)$ where $n = 5$.

$$H_0 \ : \ \theta \leq 0.5$$

$$H_1 \ : \ \theta > 0.5$$

Test 1 rejects H_0 if and only if all "success" are observed. i.e.

$$\begin{aligned} R &= \{\mathbf{x} : \mathbf{x} = (1,1,1,1,1)\} \\ &= \{\mathbf{x} : \sum_{i=1}^{5} x_i = 5\} \end{aligned}$$

1. Compute the power function
2. What is the maximum probability of making type I error?
3. What is the probability of making type II error if $\theta = 2/3$?

Solution for Test 1

$$\begin{aligned} \beta(\theta) &= \Pr(\text{reject } H_0|\theta) = \Pr(\mathbf{X} \in R|\theta) \\ &= \Pr\left(\sum X_i = 5|\theta\right) \end{aligned}$$

Because $\sum X_i \sim Binomial(5, \theta)$, $\beta(\theta) = \theta^5$.

Maximum type I error

When $\theta \in \Omega_0 = (0, 0.5]$, the power function $\beta(\theta)$ is Type I error.

$$\max_{\theta \in (0,0.5]} \beta(\theta) = \max_{\theta \in (0,0.5]} \theta^5 = 0.5^5 = 1/32 \approx 0.031$$

Type II error when $\theta = 2/3$

$$1 - \beta(\theta)|_{\theta=\frac{2}{3}} = 1 - \theta^5\Big|_{\theta=\frac{2}{3}} = 1 - (2/3)^5 = 211/243 \approx 0.868$$

Another Example

Example 2: $X_1, X_2, \cdots, X_n$ i.i.d. $Bernoulli(\theta)$ where $n = 5$.

$$H_0 \ : \ \theta \leq 0.5$$

$$H_1 \ : \ \theta > 0.5$$

Test 2 rejects H_0 if and only if 3 or more "success" are observed. i.e.

$$R = \{\mathbf{x} : \sum_{i=1}^{5} x_i \geq 3\}$$

1. Compute the power function
2. What is the maximum probability of making type I error?
3. What is the probability of making type II error if $\theta = 2/3$?

Solution for Test 2

Power function

$$\begin{aligned}\beta(\theta) &= \Pr(\sum X_i \geq 3|\theta) = \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta) + \binom{5}{5}\theta^5 \\ &= \theta^3(6\theta^2 - 15\theta + 10)\end{aligned}$$

Maximum type I error

We need to find the maximum of $\beta(\theta)$ for $\theta \in \Omega_0 = (0, 0.5]$

$$\beta'(\theta) = 30\theta^2(\theta-1)^2 > 0$$

$\beta(\theta)$ is increasing in $\theta \in (0,1)$. Maximum type I error is $\beta(0.5) = 0.5$

Type II error when $\theta = 2/3$

$$1 - \beta(\theta)|_{\theta=\frac{2}{3}} = 1 - \theta^3(6\theta^2 - 15\theta + 10)\Big|_{\theta=\frac{2}{3}} \approx 0.21$$

Sizes and Levels of Tests

Size α test

A test with power function $\beta(\theta)$ is a size α test if

$$\sup_{\theta\in\Omega_0} \beta(\theta) = \alpha$$

In other words, the maximum probability of making a type I error is α.

Level α test

A test with power function $\beta(\theta)$ is a level α test if

$$\sup_{\theta \in \Omega_0} \beta(\theta) \leq \alpha$$

In other words, the maximum probability of making a type I error is equal or less than α.

Any size α test is also a level α test

Revisiting Previous Examples
Test 1

$$\sup_{\theta \in \Omega_0} \beta(\theta) = \sup_{\theta \in \Omega_0} \theta^5 = 0.5^5 = 0.03125$$

The size is 0.03125, and this is a level 0.05 test, or a level 0.1 test, but not a level 0.01 test.

Test 2

$$\sup_{\theta \in \Omega_0} \beta(\theta) = 0.5$$

The size is 0.5

Constructing a good test

1. Construct all the level α test.
2. Within this level of tests, we search for the test with Type II error probability as small as possible; equivalently, we want the test with the largest power if $\theta \in \Omega_0^c$.

Review on standard normal and t distribution

Quantile of standard normal distribution

Let $Z \sim \mathcal{N}(0,1)$ with pdf $f_Z(z)$ and cdf $F_Z(z)$. The α-th quantile z_α or $(1-\alpha)$-th quantile $z_{1-\alpha}$ of the standard distribution satisfy

$$\Pr(Z \geq z_\alpha) = \alpha \quad \text{or} \quad z_\alpha = F_Z^{-1}(1-\alpha)$$

$$\Pr(Z \leq z_{1-\alpha}) = \alpha \quad \text{or} \quad z_{1-\alpha} = F_Z^{-1}(\alpha)$$

$$z_{1-\alpha} \;=\; -z_\alpha$$

Quantile of t distribution

Let $T \sim t_{n-1}$ with pdf $f_{T,n-1}(t)$ and cdf $F_{T,n-1}(t)$. The α-th quantile $t_{n-1,\alpha}$ or $(1-\alpha)$-th quantile $t_{n-1,1-\alpha}$ of the standard distribution satisfy

$$\Pr(T \geq t_{n-1,\alpha}) = \alpha \quad \text{or} \quad t_{n-1\alpha} = F_{T,n-1}^{-1}(1-\alpha)$$

$$\Pr(T \leq t_{n-1,1-\alpha}) = \alpha \quad \text{or} \quad t_{n-1,1-\alpha} = F_{T,n-1}^{-1}(\alpha)$$

$$t_{n-1,1-\alpha} \;=\; -t_{n-1,\alpha}$$

Likelihood Ratio Tests (LRT)

Definition Let $L(\theta|\mathbf{x})$ be the likelihood function of θ. The likelihood ratio test statistic for testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Omega} L(\theta|\mathbf{x})} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

where $\hat{\theta}$ is the MLE of θ over $\theta \in \Omega$, and $\hat{\theta}_0$ is the MLE of θ over $\theta \in \Omega_0$ (restricted MLE).

The *likelihood ratio test* is a test that rejects H_0 if and only if $\lambda(\mathbf{x}) \leq c$ where $0 \leq c \leq 1$.

Example of LRT

Example 3: Consider $X_1, \cdots, X_n$ iid $\mathcal{N}(\theta, \sigma^2)$ where σ^2 is known.

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$

Find the LRT test and its power function

Solution:

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \theta)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\sigma^2}\right] \end{aligned}$$

We need to find MLE of θ over $\Omega = (-\infty, \infty)$ and $\Omega_0 = (-\infty, \theta_0]$.

MLE of θ over $\Omega = (-\infty, \infty)$

To maximize $L(\theta|\mathbf{x})$, we need to maximize $\exp\left[-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right]$, or equivalently to minimize $\sum_{i=1}^{n}(x_i-\theta)^2$.

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\theta)^2 &= \sum_{i=1}^{n}(x_i^2+\theta^2-2\theta x_i) \\ &= n\theta^2 - 2\theta\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}x_i^2\end{aligned}$$

The equation above is minimized when $\theta = \hat{\theta} = \frac{\sum_{i=1}^{n}x_i}{n} = \overline{x}$.

- $L(\theta|\mathbf{x})$ is maximized at $\theta = \frac{\sum_{i=1}^{n}x_i}{n} = \overline{x}$ if $\overline{x} \leq \theta_0$.
- However, if $\overline{x} \geq \theta_0$, $\overline{x}$ does not fall into a valid range of $\hat{\theta}_0$, and $\theta \leq \theta_0$, the likelihood function will be an increasing function. Therefore $\hat{\theta}_0 = \theta_0$.

To summarize,

$$\hat{\theta}_0 = \begin{cases} \overline{X} & \text{if } \overline{X} \leq \theta_0 \\ \theta_0 & \text{if } \overline{X} > \theta_0 \end{cases}$$

$$\begin{aligned}\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})} &= \begin{cases} 1 & \text{if } \overline{X} \leq \theta_0 \\ \frac{\exp\left[-\frac{\sum_{i=1}^{n}(x_i-\theta_0)^2}{2\sigma^2}\right]}{\exp\left[-\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{2\sigma^2}\right]} & \text{if } \overline{X} > \theta_0 \end{cases} \\ &= \begin{cases} 1 & \text{if } \overline{X} \leq \theta_0 \\ \exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] & \text{if } \overline{X} > \theta_0 \end{cases}\end{aligned}$$

Therefore, the likelihood test rejects the null hypothesis if and only if

$$\exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] \leq c$$

and $\overline{x} \geq \theta_0$.

Specifying c

$$\begin{aligned}
\exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] &\leq c \\
\Longleftrightarrow -\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2} &\leq \log c \\
\Longleftrightarrow (\overline{x}-\theta_0)^2 &\geq -\frac{2\sigma^2 \log c}{n} \\
\Longleftrightarrow \overline{x}-\theta_0 &\geq \sqrt{-\frac{2\sigma^2 \log c}{n}} \qquad (\because \overline{x} > \theta_0)
\end{aligned}$$

So, LRT rejects H_0 if and only if

$$\begin{aligned}
\overline{x}-\theta_0 &\geq \sqrt{-\frac{2\sigma^2 \log c}{n}} \\
\Longleftrightarrow \frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} &\geq \frac{\sqrt{-\frac{2\sigma^2 \log c}{n}}}{\sigma/\sqrt{n}} = c^*
\end{aligned}$$

Therefore, the rejection region is

$$\left\{\mathbf{x} : \frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right\}$$

Power function

$$\begin{aligned}
\beta(\theta) &= \Pr(\text{reject } H_0) = \Pr\left(\frac{\overline{X}-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right) \\
&= \Pr\left(\frac{\overline{X}-\theta+\theta-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right) \\
&= \Pr\left(\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \geq \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + c^*\right)
\end{aligned}$$

Since $X_1, \cdots, X_n$ i.i.d. $\mathcal{N}(\theta, \sigma^2)$, $\overline{X} \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$. Therefore,

$$\begin{aligned}
\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} &\sim \mathcal{N}(0,1) \\
\Longrightarrow \beta(\theta) &= \Pr\left(Z \geq \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} + c^*\right)
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$.

Making size α LRT
To make a size α test,

$$\begin{aligned}
\sup_{\theta \in \Omega_0} \beta(\theta) &= \alpha \\
\sup_{\theta \leq \theta_0} \Pr\left(Z \geq \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} + c^*\right) &= \alpha \\
\Pr\left(Z \geq c^*\right) &= \alpha \\
c^* &= z_\alpha
\end{aligned}$$

Note that $\Pr\left(Z \geq \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} + c^*\right)$ is maximized when θ is maximum (i.e. $\theta = \theta_0$).

Therefore, size α LRT test rejects H_0 if and only if $\frac{\overline{x} - \theta_0}{\sigma/\sqrt{n}} \geq z_\alpha$.

Another Example of LRT

Example 4: Let $X_1, \cdots, X_n$ i.i.d. from $f(x|\theta) = e^{-(x-\theta)}$ where $x \geq \theta$ and $-\infty < \theta < \infty$. Find a LRT testing the following one-sided hypothesis.

$$
\begin{aligned}
H_0 &: \theta \leq \theta_0 \\
H_1 &: \theta > \theta_0
\end{aligned}
$$

Solution:

$$
\begin{aligned}
L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} e^{-(x_i-\theta)} I(x_i \geq \theta) \\
&= e^{-\sum x_i + n\theta} I(\theta \leq x_{(1)})
\end{aligned}
$$

The likelihood function is a increasing function of θ, bounded by $\theta \leq x_{(1)}$. Therefore, when $\theta \in \Omega = \mathbb{R}$, $L(\theta|\mathbf{x})$ is maximized when $\theta = \hat{\theta} = x_{(1)}$.

When $\theta \in \Omega_0^c$, the likelihood is still an increasing function, but bounded by $\theta \leq \min(x_{(1)}, \theta_0)$. Therefore, the likelihood is maximized when $\theta = \hat{\theta}_0 = \min(x_{(1)}, \theta_0)$. The likelihood ratio test statistic is

$$
\begin{aligned}
\lambda(\mathbf{x}) &= \begin{cases} \frac{e^{-\sum x_i + n\theta_0}}{e^{-\sum x_i + n x_{(1)}}} & \text{if } \theta_0 < x_{(1)} \\ 1 & \text{if } \theta_0 \geq x_{(1)} \end{cases} \\
&= \begin{cases} e^{n(\theta_0 - x_{(1)})} & \text{if } \theta_0 < x_{(1)} \\ 1 & \text{if } \theta_0 \geq x_{(1)} \end{cases}
\end{aligned}
$$

The LRT rejects H_0 if and only if

$$
\begin{aligned}
e^{n(\theta_0 - x_{(1)})} &\leq c \qquad (\text{and } \theta_0 < x_{(1)}) \\
\Longleftrightarrow \theta_0 - x_{(1)} &\leq \frac{\log c}{n} \Longleftrightarrow x_{(1)} \geq \theta_0 - \frac{\log c}{n}
\end{aligned}
$$

So, LRT reject H_0 is $x_{(1)} \geq \theta_0 - \frac{\log c}{n}$ and $x_{(1)} > \theta_0$. The power function is

$$\begin{aligned}
\beta(\theta) &= \Pr\left(X_{(1)} \geq \theta_0 - \frac{\log c}{n} \wedge X_{(1)} > \theta_0\right) \\
&= \Pr\left(X_{(1)} \geq \theta_0 - \frac{\log c}{n}\right)
\end{aligned}$$

To find size α test, we need to find c satisfying the condition

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \alpha$$

Constructing size α test

$$\begin{aligned}
\beta(\theta) &= \Pr\left(X_{(1)} \geq \theta_0 - \frac{\log c}{n}\right) = \prod_{i=1}^{n} \Pr\left(X_i \geq \theta_0 - \frac{\log c}{n}\right) \\
&= \prod_{i=1}^{n} \Pr\left(X_i - \theta \geq \theta_0 - \theta - \frac{\log c}{n}\right) \\
&= \prod_{i=1}^{n} \exp\left[-\theta_0 + \theta + \frac{\log c}{n}\right] = \left[\exp\left(-\theta_0 + \theta + \frac{\log c}{n}\right)\right]^n
\end{aligned}$$

which is increasing in θ. Hence

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \left[\exp\left(\frac{\log c}{n}\right)\right]^n = \alpha$$

Therefore, $\frac{\log c}{n} = \frac{1}{n}\log\alpha$, and the rejection region of the size α test is

$$R = \left\{\mathbf{X} : X_{(1)} \geq \theta_0 - \tfrac{1}{n}\log\alpha\right\}$$