I. Testing Hypotheses Problems

1. General Concepts

A. The testing problems

Q: Average GPA of class 2006 is 3.6? 

\[ H_0: \theta = 3.6 \quad \text{vs.} \quad H_1: \theta \neq 3.6 \]

(Based on a sample of 100 students.)

General Set-up.

\[ H_0: \theta \in \mathbb{R}_0 \quad \text{— Null Hypothesis} \]

v.s.

\[ H_1: \theta \in \mathbb{R}_1 = \mathbb{R} \setminus \mathbb{R}_0 \quad \text{— Alternative Hypothesis} \]

(\( \mathbb{R} \) all possible choices of \( \theta \) — Parameter Space)

(\( \mathbb{R}_0 \) a specific region of \( \mathbb{R} \))

(\( \mathbb{R}_1 \) is a complementary set of \( \mathbb{R}_0 \)).
Some terminologies:

Simple hypothesis: e.g. \( H_0: \theta = 3.6 \) or \( H_1: \theta = 2.5 \).
(contains only one value)

One-sided hypothesis: e.g. \( H_0: \theta \leq \theta_0 \) or \( H_0: \theta \geq \theta_0 \)
or \( H_1: \theta \leq \theta_0 \) etc.

two-sided hypothesis: e.g. \( H_1: \theta \neq \theta_0 \) or \( H_0: \theta \neq 2.5 \)
etc.

Two possible decisions: \( \{ \) Reject \( H_0 \) (accept \( H_1 \)) \( \} \)
\( \{ \) Not reject \( H_0 \) \( \} \)

B. Statistical Framework for testing hypotheses.
(Neyman - Pearson Paradigm)

Make decision based on sample observations:

Sample space \( \mathcal{S} \) (all possible choices of \( X = (x_1, \ldots, x_n) \))

Reject \( H_0 \) (critical region)
Accept \( H_0 \) (acceptance region)
Most times, we make decisions based on a statistic $T$.

The typical critical region is of the form:

$\{ (x_1, \ldots, x_n) \mid T > c \} \uparrow \text{known constant}$

Example: $x_1, \ldots, x_n \sim N(\mu, 1)$.

$H_0: \mu = 3.0 \quad \text{vs.} \quad H_1: \mu \neq 3.0$

Test is often based on $\bar{x}$: reject $H_0$ if

$T = |\bar{x} - 3.0| \text{ is large. (How large? answer later.)}$

So, we reject $H_0$ if

$T = |\bar{x} - 3.0| > c$ \hspace{1cm} (say $1.96/\sqrt{n}$)

**Testing procedure:** A procedure for deciding whether accept $H_0$ or not accept $H_0$ is called a testing procedure (denoted by $\delta$).
Two types of Errors

Type I error: Reject $H_0$ when the truth is $\theta \in \mathcal{R}_0$.

Type II error: fail to reject $H_0$ when the truth is $\theta \in \mathcal{R}_1 = \mathcal{R} \setminus \mathcal{R}_0$.

Probabilities of committing such errors are respectively

$$P(\mathbf{X} \in \mathcal{C} \mid \theta \in \mathcal{R}_0) \equiv P_{H_0}(\mathbf{X} \in \mathcal{C})$$

and

$$P(\mathbf{X} \in \mathcal{S} \setminus \mathcal{C} \mid \theta \in \mathcal{R}_1) = 1 - P(\mathbf{X} \in \mathcal{C} \mid \theta \in \mathcal{R}_1)$$

$$\equiv 1 - P_{H_0}(\mathbf{X} \in \mathcal{C})$$

For any $\theta \in \mathcal{R}$, we define a power function

$$\eta(\theta \mid \delta) = P(\mathbf{X} \in \mathcal{C} \mid \theta)$$

In terms of power function, the above two probabilities of Type I & II errors are $\eta(\theta \mid \delta)$ and $1 - \eta(\theta \mid \delta)$ respectively.

(Note: $H_0$ and $H_1$ are treated asymmetrically. In NP's minds, Type I error is more serious, so we need to control them.)
Ideal: we are able to minimize both types of error.

Reality: this is not possible!!

Example: \( X_1, X_2, \ldots, X_9 \sim N(\mu, 1) \). \( n = 9 \).

Ho: \( \mu = 10 \) vs \( H_1: \mu \neq 10 \).

Use \( \bar{X} \) as the testing statistic. \( \bar{X} \sim N(\mu, \frac{1}{n}) \).

Critical region: \( |\bar{X} - 10| > c \) for some \( c > 0 \)

(Say \( 5c \) )

The testing procedure depend on the choice of \( c \).

\[
\Pi(\mu|5c) = \mathcal{P}(|\bar{X} - 10| > c) = 1 - \mathcal{P}(|\bar{X} - 10| < c)
\]

\[
= 1 - \mathcal{P} \left( \frac{10 - \mu}{\sqrt{\frac{1}{n}}} \leq \frac{\bar{X} - \mu}{\sqrt{\frac{1}{n}}} \leq \frac{10 + c - \mu}{\sqrt{\frac{1}{n}}} \right)
\]

\[
= 1 - \Phi \left( \frac{30 + 3c - 3\mu}{\sqrt{3}} \right) + \Phi \left( \frac{30 - 3c - 3\mu}{\sqrt{3}} \right)
\]

If true \( \mu = 10 \), (i.e. Ho is true)

\[
P(\text{Type I error}) = \Pi(\mu = 10|5c) = 1 - \Phi(30 - 3c - 3\mu) + \Phi(30 + 3\mu + 3c)
\]

\[
= 1 - \Phi(3c) + \Phi(-3c) = 2 \left[ 1 - \Phi(3c) \right] > 0
\]

no matter which \( c \) is chosen.
If the true $\mu \neq 0$ (i.e. $H_0$ is false),

$$P(\text{Committing Type II error}) = 1 - P(\mu \neq 0 | \delta_c)$$

$$= \Phi(3c - 3(\mu-\theta)) - \Phi(-3c - 3(\mu-\theta)) \geq 0$$

depends on $\delta_c$ and $\mu$

(e.g. $\mu = 20$ (true))

$$P(\text{Committing Type II error}) = \Phi(3c - 30) - \Phi(-3c - 30) > 0$$

In extreme cases: (i) $c = 0 \Rightarrow$ reject $H_0$

- Type I error $= 1$
- Type I error $= 0$

(ii) $c = +\infty \Rightarrow$ Not reject $H_0$

- Type II error $= 1$
- Type I error $= 0$


Control Type I error and minimize type II error.

Smaller $c \Rightarrow$ smaller Type I error.
\[
\text{set } \pi(\theta|\delta) \leq \alpha \text{ for } \forall \theta \in \Theta
\]
and minimize \(1 - \pi(\theta|\delta)\) for any \(\theta \in \Theta_0\).

Such \(\alpha_0\) is called the level of significance and such a test is called a level \(\alpha_0\) test.

\[
\text{Also, we define the size of the test:}
\]
\[
\alpha(\delta) = \sup_{\theta \in \Theta_0} \pi(\theta|\delta).
\]

\[
\text{we know that for a size of } \alpha_0 \text{ test } \delta:
\]
\[
\alpha(\delta) \leq \alpha_0.
\]

C. \(p\)-values

In many fields, it is standard practice to report "p-value" instead of the testing results such as "reject/accept \(H_0\)" at \(\alpha\)-level.

Reasons:
1. Do not need to choose \(\alpha\) in advance.
2. Carries a little more information than just rejecting the results.
Example: (8.1.3) — slightly modified.

\[ X_1, \ldots, X_n \sim \text{Uniform}(0, \theta) \]

\[ H_0: \ 3 \leq \theta \leq 4 \quad \text{vs} \quad H_1: \ \theta < 3 \text{ or } \theta > 4. \]

\[ \hat{\theta}_{\text{MLE}} = Y_n = \max \{ X_1, \ldots, X_n \}^3 \]  (Example 6.5.15)

Reject \( H_0 \), if \( Y_n \geq c_1 \), or \( Y_n \leq c_2 \)

for some \( c_1 \geq 4 \)

or \( c_2 \leq 3 \)

The power function is

\[ \Pi(\theta | \delta) = P_\theta \left( Y_n \geq c_1 \quad \text{or} \quad Y_n \leq c_2 \right) \]

\[ = P_\theta \left( Y_n \geq c_1 \right) + P \left( Y_n \leq c_2 \right) \]

\[ = \begin{cases} 
1 + 0 = 1 & \text{if } \theta \leq c_2 \\
\left( \frac{c_2}{\theta} \right)^n + 0 = \left( \frac{c_2}{\theta} \right)^n & \text{if } c_2 \leq \theta \leq c_1 \\
\left( \frac{c_2}{\theta} \right)^n + 1 - \left( \frac{c_1}{\theta} \right)^n & \text{if } \theta > c_1 
\end{cases} \]

Under \( H_0: \ \theta \in [3, 4] \) \( \Delta \sim \mathcal{N} \)

so, the Type 1 error is

\[ \Pi(\theta \in [3, 4] | \delta) = \left( \frac{c_2}{\theta} \right)^n \]
So, the size of the test is

\[ \alpha(S) = \sup_{\theta \in \Theta_0} \Pi(\theta | S) = \sup_{\theta \in \Theta_0} \left( \frac{c_2}{\theta} \right)^n \]

For an \( \alpha_0 \)-test, we set

\[ \alpha(S) \leq \alpha_0 \]

i.e.

\[ \sup_{\theta \in \Theta_0} \left( \frac{c_2}{\theta} \right)^n \leq \alpha_0 \]

for all \( \theta \in \Theta_0 \)

So,

\[ c_2 = \frac{3}{\alpha_0^\frac{1}{n}} \]

So, \( c_2 = 3 \alpha_0^{\frac{1}{n}} \)

For an \( \alpha_0 \)-level test, we want to know the type II error.

Under \( H_1 \), \( \theta < 3 \) or \( \theta > 4 \).

(\( \theta \notin \Theta_0 \))

So, the type II error is

\[ 1 - \Pi(\theta \notin \Theta_0 | S) = \begin{cases} 
0 & \text{if } \theta \leq 3\alpha_0^\frac{1}{n} \\
\frac{3^n \alpha_0}{\theta^n} & \text{if } 3\alpha_0^\frac{1}{n} < \theta < 3 \\
\frac{3^n \alpha_0}{\theta^n} & \text{if } 4 \leq \theta < 4_1 \\
\frac{3^n \alpha_0}{\theta^n} + 1 - \frac{c_1}{\theta^n} & \text{if } \theta \geq 4_1 
\end{cases} \]

If we pick \( C_1 = 4 \), the type II error is minimized.

\[ \Rightarrow \text{The test to use: } \sqrt{n} Y_n \geq 3 \alpha_0^\frac{1}{n} \text{ or } Y_n \leq 4. \]