MH1820 Week 12

Recall the following ...

Procedue for Hypothesis Testing:

- Given are observations x_1, ..., x_n.
- Formulate **null hypothesis** H_0 describing the population distribution from which observations were drawn.
- Choose significance level α (often $\alpha = 0.05$)
- Choose test statistic $T(X_1, \ldots, X_n)$ that contains information on the parameters involved in H_0 and whose distribution is known under H_0.
- Assuming H_0, compute probability (p-value) to observe $t = T(x_1, \ldots, x_n)$ or something "**at least as extreme as t" (in the direction of rejection of H_0)**.
- If the p-value is smaller than α, reject null hypothesis.

- In the p-value approach, we reject H_0 when p-value is less than the significance level α.
- Instead of using p-value, we can also formulate rejection criteria using **rejection region**, where we reject H_0 if the test statistic statisfies certain inequalities.

  E.g. Reject $H_0 \iff \sum_{i=1}^{n} X_i > c$, $|\sum_{i=1}^{n} X_i| > c$ etc.

There are two types of errors in hypothesis testing:

H_0	H_0 True	H_0 False
Reject H_0	Type I Error	
Do not Reject H_0		Type II Error

Type I Errors and Size of a Test

Size of a Test

- If the null hypothesis H_0 is true, but **rejected**, then a **Type I Error** occurs.
- The probability of a Type I Error is

$$\mathbb{P}(H_0 \text{ rejected} | H_0).$$

- $\mathbb{P}(H_0 \text{ rejected} | H_0)$ is also called the **size** of the test.
- The smaller the size, the more conclusive is the test – the size measures how conclusive a test is.

Example 1

- X_1, …, X_9 i.i.d $\sim N(\mu, 1)$
- Null hypothesis $H_0 : \mu = 0$
- Rejection criteria: Reject $H_0 \iff |\sum_{i=1}^{9} X_i| > 5.88$.

Compute the size of the test.

Solution. The size of the test is

$$\mathbb{P}(H_0 \text{ rejected}|H_0) = \mathbb{P}(|\sum_{i=1}^{9} X_i| > 5.88|H_0).$$

Assuming H_0 is true, the standardized sample mean is standard normal:

$$\frac{\frac{\sum_{i=1}^{9} X_i}{9} - 0}{1/\sqrt{9}} = \frac{\sum_{i=1}^{9} X_i}{3} \sim N(0,1).$$

$$
\begin{aligned}
\mathbb{P}(|\sum_{i=1}^{9} X_i| > 5.88 | H_0) &= \mathbb{P}(\frac{|\sum_{i=1}^{9} X_i|}{3} > 1.96 | H_0) \\
&= 2\Phi(-1.96) \approx 0.05.
\end{aligned}
$$

$$
\implies \text{size of the test} = 0.05.
$$

$\square$

Example 2

- $X_1, \ldots, X_{100}$ i.i.d $\sim \mathit{Bernoulli}(p)$, $0 \leq p \leq 1$
- Null hypothesis $H_0 : p = 0.5$
- Test statistic $T = X_1 + \cdots + X_{100}$
- Rejection criteria: Reject $H_0 \iff T - 50 > 8$.

Compute the size of the test.

Solution. Assuming H_0 is true, we have $T \sim \mathit{Binomial}(100, p = 0.5)$

$$
\begin{aligned}
\mathbb{P}(H_0 \text{ rejected} | H_0) = \quad &= \mathbb{P}(T - 50 > 8 | p = 0.5) \\
&= \mathbb{P}(T > 58 | p = 0.5) \\
&= 1 - \mathbb{P}(T \leq 58 | p = 0.5) \\
&\approx 1 - \Phi\left(\frac{58 - 100(0.5)}{0.5\sqrt{100}}\right) \quad \text{by CLT} \\
&\approx 0.05.
\end{aligned}
$$

$$\Longrightarrow \text{ size of the test} = 0.0548.$$

Example 3

- $X_1, \ldots, X_4$ i.i.d $\sim Bernoulli(p)$, $0 \leq p \leq 1$
- Null hypothesis $H_0 : p = 0.5$
- Test statistic $T = X_1 + X_2 + X_3 + X_4$
- Rejection criteria: Reject $H_0 \iff |T - 2| \geq 2$.

Compute the size of the test.

Solution. Assuming H_0 is true, we have

$$T \sim Binomial(4, 0.5).$$

$$\begin{aligned}
\mathbb{P}(H_0 \text{ rejected} | H_0) = \quad &= \mathbb{P}(|T-2| \geq 2 | p = 0.5) \\
&= \mathbb{P}(T \leq 0 | p = 0.5) + \mathbb{P}(T \geq 4 | p = 0.5) \\
&= \mathbb{P}(T = 0 | p = 0.5) + \mathbb{P}(T = 4 | p = 0.5) \\
&= \binom{4}{0}(0.5)^0(0.5)^4 + \binom{4}{4}(0.5)^4(0.5)^0 \\
&= 0.125.
\end{aligned}$$

$$\Longrightarrow \text{ size of the test} = 0.125.$$

□

Type II Errors and Power of a Test

Type II Error

- If the null hypothesis H_0 is wrong, but **not rejected**, then a **Type II** error occurs.
- It is usually not possible to make both Type I and Type II errors arbitrarily small.
- Realistic goal: Find test with a prescribed probability of a Type I Error that minimizes the probability of a Type II Error.
- Type II Error can be controlled using the Alternative Hypothesis.

Example 4

Coin is tossed 100 times to test if there is a bias towards heads.

- $X_1, \ldots, X_{100}$ i.i.d $\sim Bernoulli(p)$
- $H_0 : p = 0.5, \quad H_1 : p > 0.5$

Type I Error: Coin is fair, but H_0 is rejected.

Type II Error: Coin is biased towards heads, but H_0 is not rejected.

Power of a Test

- The probability of a Type II Error is denoted by β:

$$\beta = \mathbb{P}(H_0 \text{ not rejected} | H_1)$$

- The probability that H_0 is rejected if it is wrong is the **power** of the test, i.e.

$$\textbf{Power} = \mathbb{P}(H_0 \text{ rejected} | H_1) = 1 - \beta.$$

Example 5

- $X_1, \ldots, X_{10}$ i.i.d $\sim Poisson(\lambda)$
- Test $H_0 : \lambda = \frac{1}{10}$ against $H_1 : \lambda = 1$
- H_0 is rejected $\Longleftrightarrow \sum_{i=1}^{10} X_i \geq 2$.

Find the size and power of this test.

Solution.
By checking MGF, it is readily verified that (Tutorial 6 Problem 3)

$$\sum_{i=1}^{10} X_i \sim Poisson(10\lambda).$$

Assuming H_0 is true, then $\lambda = \frac{1}{10} = 0.1$, and so $\sum_{i=1}^{10} X_i \sim Poisson(10\lambda) = Poisson(1)$. Thus,

$$\begin{aligned}
\text{Size of test} &= \mathbb{P}(H_0 \text{ rejected} | H_0) \\
&= \mathbb{P}(\sum_{i=1}^{10} X_i \geq 2 | \lambda = 0.1) \\
&= 1 - \mathbb{P}(\sum_{i=1}^{10} X_i \leq 1 | \lambda = 0.1) \\
&= 1 - \left(e^{-1} \frac{1^0}{0!} + e^{-1} \frac{1^1}{1!} \right) \\
&= 1 - e^{-1} - e^{-1} \approx 0.26.
\end{aligned}$$

Assuming H_1 is true, $\lambda = 1$, and so $\sum_{i=1}^{10} X_i \sim Poisson(10\lambda) = Poisson(10)$. Thus,

$$\begin{aligned}
\beta &= \mathbb{P}(H_0 \text{ NOT rejected} | H_1) \\
&= \mathbb{P}(\sum_{i=1}^{10} X_i \leq 1 | \lambda = 1) \\
&= e^{-10}\frac{10^0}{0!} + e^{-10}\frac{10^1}{1!} \\
&= 11e^{-10} \approx 0.0005.
\end{aligned}$$

$$\implies \text{Power of test} = 1 - \beta \approx 0.9995.$$

□

Example 6

- $X_1, \ldots, X_{10}$ i.i.d $\sim Poisson(\lambda)$
- Test $H_0 : \lambda = \frac{1}{10}$ against $H_1 : \lambda = 1$
- Test statistic: $\sum_{i=1}^{10} X_i \sim Poisson(10\lambda)$
- Rejection criteria: Reject $H_0 \iff \mathbb{P}(\sum_{i=1}^{10} X_i > c)$.

Suppose we require the size of the test to be at most 0.05. What is the maximum power we can achieve?

Solution.

$$\begin{aligned} \text{Size} &= \mathbb{P}(H_0 \text{ rejected} | H_0) \\ &= \mathbb{P}(\sum_{i=1}^{10} X_i > c | \lambda = 0.1) \\ &= 1 - F(c), \end{aligned}$$

where F is the CDF of *Poisson*(1).

From Table: $F(2) \approx 0.92$, $F(3) \approx 0.981$. So

$$\text{Size} \leq 0.05 \implies 1 - F(c) \leq 0.05 \implies F(c) \geq 0.95 \implies c \geq 3.$$

We have shown that $c \geq 3$ if the test has size ≤ 0.05.

$$
\begin{aligned}
\text{Power} &= \mathbb{P}(H_0 \text{ rejected} | H_1) \\
&= \mathbb{P}(\sum_{i=1}^{10} X_i > c | \lambda = 1) \\
&= 1 - G(c) \quad (\text{where } G \text{ is the CDF of } \textit{Poisson}(10)) \\
&\leq 1 - G(3) \quad (\text{since } c \geq 3) \\
&\approx 0.99.
\end{aligned}
$$

With size of the test ≤ 0.05, the maximum power the test can achieve based on given rejection criteria is 0.99 (is attained if we set $c = 3$). $\square$.

Example 7

- $X_1, \ldots, X_{25}$ i.i.d $\sim N(\mu, 100)$.
- Test $H_0 : \mu = 60$ against $H_1 : \mu > 60$
- Reject $H_0 \iff \overline{X} \geq c$.

Compute the size of the test for $c = 62$ and $c = 63.29$. For each of the c above, what is the power of the test if $\mu = 65$?

Solution.

Assuming H_0 is true, i.e $\mu = 60$. Note: the standardized sample mean is standard normal:

$$\frac{\overline{X} - 60}{10/\sqrt{25}} = \frac{\overline{X} - 60}{2} \sim N(0, 1).$$

- If $c = 62$, then

$$\text{Size} = \mathbb{P}\left(\frac{\overline{X}-60}{2} \geq \frac{62-60}{2}\right) = 1-\Phi(1) = 1-0.8413 = 0.1587$$

- If $c = 63.29$, then

$$\text{Size} = \mathbb{P}\left(\frac{\overline{X}-60}{2} \geq \frac{63.29-60}{2}\right) = 1-\Phi(1.645) = 0.05$$

Assuming H_1 holds with $\mu = 65$.
We have

$$\beta = \mathbb{P}(H_0 \text{ not rejected} | H_1) = \mathbb{P}\left(\frac{\overline{X} - 65}{2} < \frac{c - 65}{2}\right) = \Phi\left(\frac{c - 65}{2}\right)$$

- $c = 62 \implies \text{Power} = 1 - \beta = 1 - \Phi(-1.5) = \Phi(1.5) = 0.9332$
- $c = 63.29 \implies \text{Power} = 1 - \beta = 1 - \Phi(-0.855) \approx \Phi(0.86) = 0.8051$

$\square$

Remark: Though the size of the test decreases as we increase c from 62 to 63.29, the power of the test reduces from 0.9332 to 0.8051.