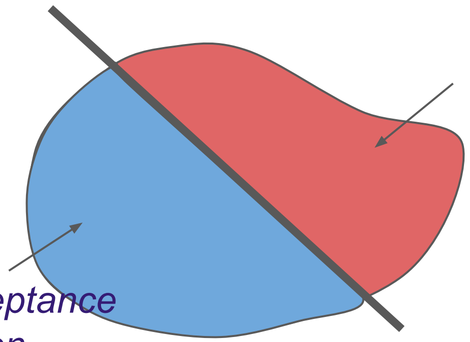
Type I *vs* Type II Error

9.2 of Rice

07/07/2021

Berkeley
UNIVERSITY OF CALIFORNIA

In the previous lecture,



- Exact sampling distributions under $N(\mu, \sigma^2)$.
  - $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 / \sigma^2 \sim \chi^2_{n-1}$.
  - Independence between sample mean and variance.
  - Exact CIs for μ and σ^2.

- Hypothesis testing:
  - Null hypothesis versus alternative hypothesis.
  - Asymmetric nature of HT.
  - Rejection region: $R = \{\text{Unlikely } T(\mathbf{X}_n) \text{ values under } H_0\}$.
  - $\mathrm{P}(\text{Type I error})$ is the significance level.
  - $1 - \mathrm{P}(\text{Type II error})$ is the power of test.

- Likelihood ratio test (LRT)
  - $H_0 : \theta \in \Theta_0 \quad \leftrightarrow \quad H_1 : \theta \in \Theta_1$
  - Rule-based method to formulate rejection region.
  - $\lambda(\mathbf{X}_n) = \dfrac{\sup_{\Theta_0} L(\theta \,|\, \mathbf{X}_n)}{\sup_{\Theta} L(\theta \,|\, \mathbf{X}_n)}$ with $R = \{\lambda(\mathbf{X}_n) \le c\}$

Type I error is more serious

	H_0 is true	H_1 is true
Reject H_0	Type I error	Correct decision
Fail to reject H_0	Correct decision	Type II error

$$\begin{aligned} \mathrm{P}(\text{Type I error}) &= \mathrm{P}[\mathbf{X}_n \in R \,|\, \Theta_0], \\ \mathrm{P}(\text{Type II error}) &= \mathrm{P}[\mathbf{X}_n \notin R \,|\, \Theta_1]. \end{aligned}$$

⇩

The rejection region of a LRT should be

$$R = \{\lambda(\mathbf{X}_n) \le c\}$$

in which $0 \le c \le 1.$

$$\begin{aligned} \mathrm{P}(\text{Type I error}) &= \mathrm{P}[\lambda(\mathbf{X}_n) \le c \,|\, \Theta_0], \\ \mathrm{P}(\text{Type II error}) &= \mathrm{P}[\lambda(\mathbf{X}_n) > c \,|\, \Theta_1]. \end{aligned}$$

***Strategy**: Minimize type II error <u>after</u> making sure* $\mathrm{P}(\textit{type I error}) \le \alpha$.

Power function

Definition. The power function of a test with rejection region R is defined for any $\theta \in \Theta$,

$$\beta(\theta) = P(\mathbf{X}_n \in R \,|\, \theta).$$

$$\beta(\theta) = \begin{cases} \text{Type I} & , \theta \in \Theta_0 \\ \text{power} = 1 - \text{Type II} & , \theta \in \Theta_1 \end{cases}$$

The power of a test

$= 1 - P(\text{Type II error})$

$= 1 - P(X_n \notin R \,|\, \Theta_1)$

$= P(X_n \in R \,|\, \Theta_1)$

$$\text{P(Type I error)} = \text{P}[\mathbf{X}_n \in R \,|\, \Theta_0],$$
$$\text{P(Type II error)} = \text{P}[\mathbf{X}_n \notin R \,|\, \Theta_1].$$

⇓

$$\text{P(Type I error)} = \text{P}[\lambda(\mathbf{X}_n) \le c \,|\, \Theta_0],$$
$$\text{P(Type II error)} = \text{P}[\lambda(\mathbf{X}_n) > c \,|\, \Theta_1].$$

Strategy: *Minimize type II error <u>after</u> making sure* P(*type I error*) $\le \alpha$.

Power function

Given μ, $\bar{X}_n \sim N\left(\mu, \left(\frac{1}{\sqrt{n}}\right)^2\right)$

$\Phi(x) = P(Z \le x)$ where $Z \sim N(0,1)$

Example 2 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, 1)$. Consider testing

$$H_0 : \mu = \mu_0 \quad \leftrightarrow \quad H_1 : \mu \neq \mu_0.$$

composite

Find the power function of the LRT.

Solution. We know the LRT rejection region is $R = \{|\bar{X}_n - \mu_0| \ge c'\}$ with $c' = \sqrt{\frac{2 \log 1/c}{n}}$.

By definition, $\beta(\mu) = P(R \mid \mu) = P(|\bar{X}_n - \mu_0| \ge c' \mid \mu)$

$= P\left(\frac{|\bar{X}_n - \mu + \mu - \mu_0|}{1/\sqrt{n}} \ge \frac{c'}{1/\sqrt{n}} \,\middle|\, \mu\right) = P\left(\left|\underbrace{\frac{\bar{X}_n - \mu}{1/\sqrt{n}}}_{= Z \sim N(0,1)} + \frac{\mu - \mu_0}{1/\sqrt{n}}\right| \ge \frac{c'}{1/\sqrt{n}} \,\middle|\, \mu\right)$

$= P\left(\left|Z + \frac{\mu - \mu_0}{1/\sqrt{n}}\right| \ge \frac{c'}{1/\sqrt{n}}\right) = P\left(Z \ge \frac{c'}{1/\sqrt{n}} - \frac{\mu - \mu_0}{1/\sqrt{n}}\right) + P\left(Z \le -\frac{c'}{1/\sqrt{n}} - \frac{\mu - \mu_0}{1/\sqrt{n}}\right)$

$= 1 - \Phi\left(\frac{c'}{1/\sqrt{n}} - \frac{\mu - \mu_0}{1/\sqrt{n}}\right) + \Phi\left(\frac{-c'}{1/\sqrt{n}} - \frac{\mu - \mu_0}{1/\sqrt{n}}\right)$

sample size

Strategy: *Minimize type II error* *__after__* *making sure* P(*type I error*) $\le \alpha$.

z_α such that

$$P(Z > z_\alpha) = \alpha$$

★ $$P(\text{Type I error}) = P(R \mid H_0) = P(R \mid \mu = \mu_0)$$

$$= \beta(\mu_0) = P\left(|Z| \geq \frac{c'}{1/\sqrt{n}} \right) \overset{=}{\cancel{\leq}} \alpha$$

$$\frac{c'}{1/\sqrt{n}} = z_{\alpha/2}.$$

★ $$1 - P(\text{Type II error}) = 1 - \Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu}{1/\sqrt{n}} \right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu}{1/\sqrt{n}} \right).$$

$$= \beta(\mu)$$

$$\alpha, \qquad n$$

$$\uparrow \qquad \uparrow$$

Power function

Example 2 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, 1)$. Consider testing

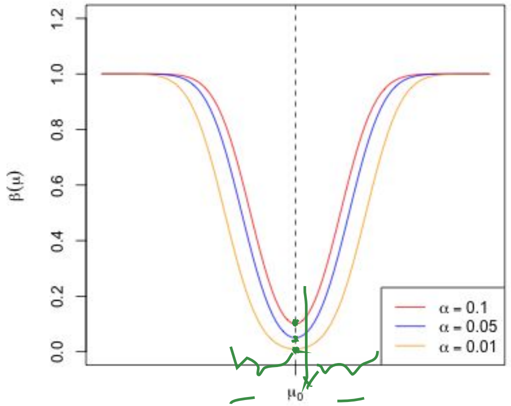
$$H_0 : \mu = \mu_0 \quad \leftrightarrow \quad H_1 : \mu \neq \mu_0.$$

Find the power function of the LRT.

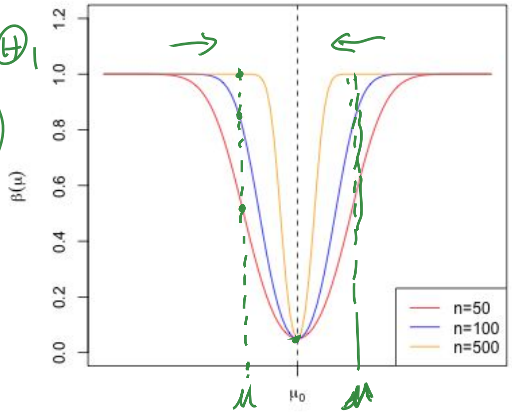
$$\beta(\theta) = \begin{cases} P(\text{Type I}) \downarrow, & \theta \in \Theta_0 \\ \text{power} \uparrow, & \theta \in \Theta_1 \end{cases}$$

$= 1 - P(\text{Type II})$

n=50 fixed



α=0.05 fixed



$n \uparrow$

power $\uparrow$

Strategy: *Maximize power after making sure* P(*type I error*) $\leq \alpha$.

= Minimize Type II error

Neyman-Pearson Lemma

Simple hypotheses testing

9.2 of Rice

07/07/2021

Likelihood ratio test ($Fix\ c$)

Under Θ_0, $\bar{X}_n - 0.49 \to N\left(0, \frac{0.49 * 0.51}{n}\right)$

Under Θ_1, $\bar{X}_n - 0.51 \to N\left(0, \frac{0.49 * 0.51}{n}\right)$

Example 3 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $\mathrm{Bernoulli}(p)$. Consider testing

$$H_0 : p = 0.49 \quad \leftrightarrow \quad H_1 : p = 0.51.$$

Approximate type I and II error using CLT.

$$\lambda(\mathbf{X}_n) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{X}_n)}{\sup_{\Theta} L(\theta \mid \mathbf{X}_n)} = \min\left\{1, \left(\frac{0.49}{0.51}\right)^{2\sum X_i - n}\right\}. \Rightarrow R = \{\lambda(\mathbf{X}_n) \le c\} = \{\bar{X}_n \ge c'\} \text{ with } c' = \frac{1}{2} + \frac{\log c}{2n \log \frac{0.49}{0.51}}.$$

P(Type I error) = $P(R \mid \Theta_0)$

$= P(\bar{X}_n \ge c' \mid p = 0.49)$

$\approx P\left(\underbrace{\frac{\sqrt{n}(\bar{X}_n - 0.49)}{\sqrt{0.49 * 0.51}}}_{= Z \sim N(0,1)} \ge \frac{\sqrt{n}(c' - 0.49)}{\sqrt{0.49 * 0.51}} \,\middle|\, p = 0.49\right)$

$= 1 - \Phi\left[\frac{\sqrt{n}(c' - 0.49)}{\sqrt{0.49 * 0.51}}\right]$

P(Type II error) = $1 - P(R \mid \Theta_1)$

$= P(\bar{X}_n < c' \mid p = 0.51)$

$= P\left(\frac{\sqrt{n}(\bar{X}_n - 0.51)}{\sqrt{0.49 * 0.51}} < \frac{\sqrt{n}(c' - 0.51)}{\sqrt{0.49 * 0.51}} \,\middle|\, p = 0.51\right)$

$= \Phi\left[\frac{\sqrt{n}(c' - 0.51)}{\sqrt{0.49 * 0.51}}\right]$

Strategy: *Minimize type II error* *<u>after</u>* *making sure* P(*type I error*) $\le \alpha$.

Likelihood ratio test

Solution cont'd.

Find n such that
$P(\textit{type I error}) \leq 0.01$,
$P(\textit{type II error}) \leq 0.01$.

To control the type I error,

$$1-\Phi\left(\frac{\sqrt{n}\,(c'-0.49)}{\sqrt{0.49*0.51}}\right) \leq \alpha = 0.01$$

$$\frac{\sqrt{n}\,(c'-0.49)}{\sqrt{0.49*0.51}} \geq z_\alpha = 2.326$$

$$\sqrt{n} \geq \frac{2.326 * \sqrt{0.49*0.51}}{c'-0.49}$$

$$= \frac{1.163}{0.01 - \frac{\log c}{n \log \frac{0.49}{0.51}}} \geq \frac{1.163}{0.01} = 116.28 \Rightarrow n \geq 13520.27$$

To control type II error,

$$\Phi\left(\frac{\sqrt{n}\,(c'-0.51)}{\sqrt{0.49*0.51}}\right) \leq \alpha = 0.01$$

under H_0, $\sum X_i \sim \text{Binom}(n, 0.49)$

Likelihood ratio test (*Fix n*)

Example 3 *cont'd*. Let $X_1, \ldots, X_n$ be i.i.d $\text{Bernoulli}(p)$. Consider testing

$$H_0 : p = 0.49 \quad \leftrightarrow \quad H_1 : p = 0.51.$$

Calculate type I and II error.

$$\lambda(\mathbf{X}_n) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{X}_n)}{\sup_{\Theta} L(\theta \mid \mathbf{X}_n)} = \min\left\{1, \left(\frac{0.49}{0.51}\right)^{2\sum X_i - n}\right\}. \Rightarrow R = \{\lambda(\mathbf{X}_n) \le c\} = \{\bar{X}_n \ge c'\} \text{ with } c' = \frac{1}{2} + \frac{\log c}{2n \log \frac{0.49}{0.51}}.$$

P(Type I error) = $P(R \mid p = 0.49)$

$= P(\bar{X}_n \ge c' \mid p = 0.49) = P\left(\sum_{i=1}^{n} X_i \ge nc' \mid p=0.49\right)$

$= \sum_{k=\lceil nc' \rceil}^{n} \binom{n}{k} 0.49^k (1-0.49)^{n-k} \le \alpha$

in which $\lceil \cdot \rceil$ is the ceiling function.

P(Type II error) = $1 - P(R \mid p = 0.51)$

$= P(\bar{X}_n < c \mid p = 0.49)$

$= \sum_{k=0}^{\lfloor nc' \rfloor} \binom{n}{k} 0.51^k \, 0.49^{n-k}$.

in which $\lfloor \cdot \rfloor$ is the floor function.

Strategy: *Minimize type II error after making sure* $\text{P}(\textit{type I error}) \le \alpha$.

Likelihood ratio test ($Fix\ n$)

Neyman-Pearson Lemma. Consider simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Among all tests that have $\mathrm{P}(type\ I\ error) \le \alpha$, LRT with $\mathrm{P}(type\ I\ error) = \alpha$ has the maximum power.

Proof*. To control Type I error, what if we make $c' \to \infty$, and

Heuristic proof using the previous example.

P(Type I error) = 0

P(Type II error) = 1.

$c' \uparrow$, P(Type I error) $\downarrow$ P(Type II error) $\uparrow$

Pick c', P(Type I error) $= \alpha$, P(Type II error) would be the minimum among all tests that has P(Type I) $\le \alpha$.

Strategy: *Minimize type II error <u>after</u> making sure* $\mathrm{P}(type\ I\ error) \le \alpha$.

Likelihood ratio test

General proof

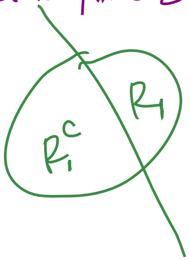
Proof cont'd.

$\beta(\theta_0) = P(R \mid \theta_0) = P(\text{Type I error})$

Θ_0 singleton

$H_0: \theta = \theta_0 \quad \text{vs} \quad H_1: \theta = \theta_1$

Need to prove LRT with $\beta(\theta_0) = \alpha$ has the maximum power among all tests with $\beta(\theta_0) \le \alpha$.

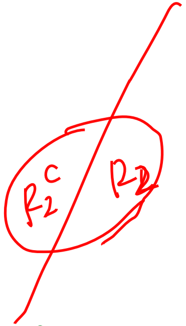


$$\delta_1(\bar{X}_n) = \begin{cases} 1, & \bar{X}_n \in R_1 \\ 0, & \bar{X}_n \notin R_1 \end{cases} = \mathbb{1}\{\bar{X}_n \in R_1\}$$

Assume that δ_1 is LRT with $\beta_1(\theta_0) = \alpha$.

$$R_1 = \{\lambda(\bar{X}_n) \le c\} = \left\{ \frac{f(\bar{X}_n \mid \theta_0)}{\max\{f(\bar{X}_n \mid \theta_0), f(\bar{X}_n \mid \theta_1)\}} \le c \right\}$$

$$= \left\{ \min\left\{1, \frac{f(\bar{X}_n \mid \theta_0)}{f(\bar{X}_n \mid \theta_1)}\right\} \le c \right\} = \left\{ \frac{f(\bar{X}_n \mid \theta_0)}{f(\bar{X}_n \mid \theta_1)} \le c \right\}.$$



Given any other test with $\beta_2(\theta_0) \le \alpha$,

$$\delta_2(\bar{X}_n) = \mathbb{1}\{\bar{X}_n \in R_2\}$$

$$0 \ge \underbrace{\overbrace{\Big[\underbrace{\overbrace{\delta_1(\bar{X}_n)}^{=0}}_{=1} - \underbrace{\overbrace{\delta_2(\bar{X}_n)}^{1 \text{ or } 0}}_{1 \text{ or } 0}\Big]}^{\le 0}}_{\ge 0} \underbrace{\overbrace{\Big[f(\bar{X}_n \mid \theta_0) - c f(\bar{X}_n \mid \theta_1)\Big]}^{>0}}_{\le 0}$$

Cont'd :

$$0 \geq \int [\delta_1(\bar{X}_n) - \delta_2(\bar{X}_n)][f(\bar{X}_n|\theta_0) - c f(\bar{X}_n|\theta_1)]\, d\bar{X}_n$$

$$= \int \delta_1(\bar{X}_n) f(\bar{X}_n|\theta_0)\, d\bar{X}_n - c\int \delta_1(\bar{X}_n) f(\bar{X}_n|\theta_1)\, d\bar{X}_n$$

$$- \int \delta_2(\bar{X}_n) f(\bar{X}_n|\theta_0)\, d\bar{X}_n + c\int \delta_2(\bar{X}_n) f(\bar{X}_n|\theta_1)\, d\bar{X}_n$$

$$= P(\bar{X}_n \in R_1 | \theta_0) - c\, P(\bar{X}_n \in R_1 | \theta_1)$$

$$- P(\bar{X}_n \in R_2 | \theta_0) + c\, P(\bar{X}_n \in R_2 | \theta_1)$$

$$= \underbrace{\beta_1(\theta_0)}^{=\alpha} - c\,\beta_1(\theta_1) - \underbrace{\beta_2(\theta_0)}_{\leq \alpha} + c\,\beta_2(\theta_1) \geq \underset{(0,1)}{c}\,(\beta_2(\theta_1) - \beta_1(\theta_1))$$

≥ 0

$$\Rightarrow \beta_2(\theta_1) \geq \beta_1(\theta_1). \quad \square$$

Pre-specify the significance level

The choice of significance level is influenced by custom. Small values like 0.01 and 0.05 are commonly used.

Example 4. Let $X_1, \ldots, X_{20}$ be i.i.d $N(\mu, 1)$. Consider testing

$$H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu = 2.$$

Find the LRT with the significance level 0.05. If $\bar{X}_{20} = 1.536$, do we reject H_0? What about $\bar{X}_{20} = 1.368$?

Solution. The LRT rejection region is $R = \{\lambda(\mathbf{X}_n) \le c\} = \{\bar{X}_n \ge c'\}$.

$$\lambda(\bar{x}_n) = \frac{\sup_{\Theta_0} L(\theta \mid \bar{x}_n)}{\sup_{\Theta} L(\theta \mid \bar{x}_n)} = \frac{L(\mu=1 \mid \bar{x}_n)}{\max\{L(\mu=1 \mid \bar{x}_n),\ L(\mu=2 \mid \bar{x}_n)\}} \le c$$

$$\Leftrightarrow \frac{L(\mu=1 \mid \bar{x}_n)}{L(\mu=2 \mid \bar{x}_n)} \le c.$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum (x_i-1)^2}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum (x_i-2)^2}} = e^{-\frac{1}{2}\left[\sum_{i=1}^n (x_i-1)^2 - \sum_{i=1}^n (x_i-2)^2\right]}$$

$$\cancel{x_i^2} + 1 - 2x_i \qquad \cancel{x_i^2} + 4 - 4x_i$$

$$= e^{-\frac{1}{2}\sum_{i=1}^n 2x_i - 3} = e^{\frac{3}{2}} e^{-\sum_{i=1}^n x_i} \le c$$

$$\boxed{\sum_{i=1}^n x_i \ge \log e^{3/2} - c}$$

$\bar{X}_n \sim N\left(1, \frac{1}{n}\right)$

To make the significance level $= \alpha$,

$$\beta(1) = P(R \mid \mu=1) = P(\bar{X}_n \geq c' \mid \mu=1) = \alpha$$

$$= P\left(\underbrace{\frac{\bar{X}_n - 1}{1/\sqrt{n}}}_{Z} \geq \frac{c'-1}{1/\sqrt{n}} \,\Big|\, \mu=1\right)$$

$$= 1 - \Phi\left(\frac{c'-1}{1/\sqrt{n}}\right) = \alpha.$$

By definition of z_α, we know that:

$$\frac{c'-1}{1/\sqrt{n}} = z_\alpha = 1.64 \quad \text{with} \quad n = 20$$

$$\Rightarrow \quad c' = \frac{1.64}{\sqrt{20}} + 1 = 1.367.$$

Therefore, LRT with the maximum power is $\delta(\bar{X}_n) = \mathbb{1}\{\bar{X}_n \geq 1.367\}$.

If $\bar{X}_n = 1.536$, we conclude that the null hypothesis is rejected.

If $\bar{X}_n = 1.368$, we conclude that the null hypothesis is rejected.

Pre-specify the significance level

The choice of significance level is influenced by custom. Small values like 0.01 and 0.05 are commonly used.

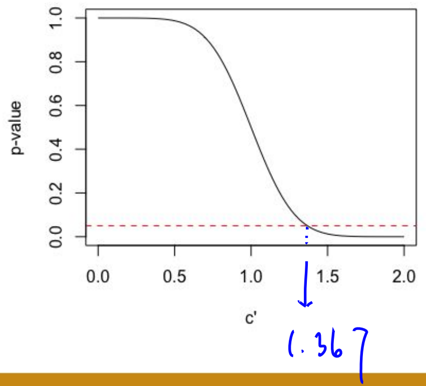
Example 4. Let $X_1, \ldots, X_{20}$ be i.i.d $N(\mu, 1)$. Consider testing

$$H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu = 2.$$

Find the LRT with the significance level 0.05. If $\bar{X}_{20} = 1.536$, do we reject H_0? What about $\bar{X}_{20} = 1.368$?

Solution cont'd. The LRT rejection region is $R = \{\lambda(\mathbf{X}_n) \leq c\} = \{\bar{X}_n \geq c'\}$.

$$\text{We want to find } c' \text{ such that } \mathrm{P}(\bar{X}_n \geq c' \mid H_0) = 0.05.$$



Which sample has stronger evidence against H_0?

p-value

Definition. Under H_0, the probability of observing a result *at least as extreme as* the test statistic.

<u>Example 4 *cont'd*</u>. $H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu = 2.$

$\bar{X}_n = 1.536$. what does it mean to be extreme for null hypothesis?

$$P(\bar{X}_n \geq 1.536 \mid H_0) = P\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \geq \frac{1.536-1}{1/\sqrt{20}}\right) = 0.00826$$

$\bar{X}_n = 1.368$

$$P(\bar{X}_n \geq 1.368 \mid H_0) = P\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \geq \frac{1.368-1}{1/\sqrt{20}}\right) = 0.0499.$$

Which sample has stronger evidence against H_0?

p-value

Sometimes a significance level cannot be attained. In this case, p-value can be used as an alternative.

$\bar{X}_n = \frac{12}{16}$

$P(\bar{X}_n \geq \frac{12}{16} \mid p=0.49) = 0.0322$

Example 3 *cont'd*. Let $X_1, \ldots, X_{16}$ be i.i.d $\text{Bernoulli}(p)$. Consider testing

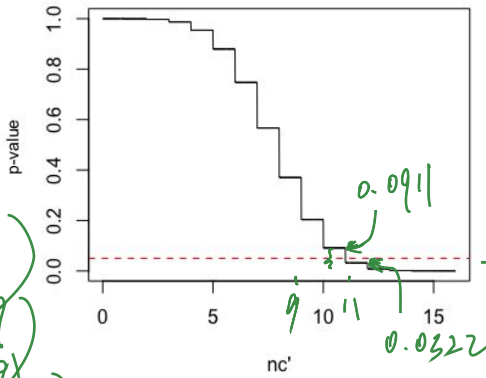
$$H_0 : p = 0.49 \quad \leftrightarrow \quad H_1 : p = 0.51.$$

Can we find the LRT with the significance level 0.05? If $\bar{X}_{16} = 11/16$, what is the p-value?

Solution. The LRT rejection region is $R = \{\lambda(\mathbf{X}_n) \leq c\} = \{\bar{X}_n \geq c'\}$.

We *cannot* find c' such that $\mathrm{P}(\bar{X}_n \geq c' \mid H_0) = 0.05$.

However, we can still calculate the p-value.



$P(\bar{X}_n \geq c' \mid p=0.49)$

$= \sum_{k=\lceil nc' \rceil}^{n} \binom{n}{k} 0.49^k \, 0.51^{n-k}$

$P(\bar{X}_n \geq \frac{11}{16} \mid p=0.49)$

$= P(\sum X_i \geq 11 \mid p=0.49)$

Binom(16, 0.49)

$= 1 - P(\sum X_i \leq 10 \mid p=0.49)$

$= 0.0911$

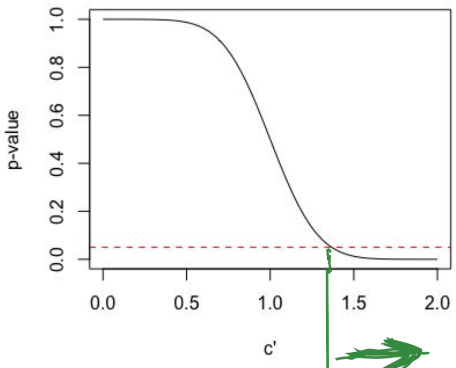
```
1 - pbinom(x, size=16, prob=0.49)
```

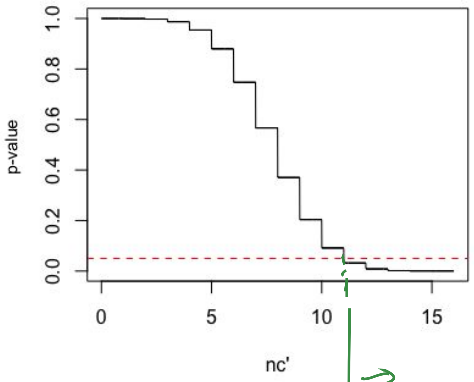
10

p-*value* approach to HT

p -value ↘ Evidence against H_0 ↗

- If $p\text{-value} > \alpha$, we fail to reject H0 and conclude there is not enough evidence against H0.
- If $p\text{-value} \leq \alpha$, we reject H0 and conclude the test results are statistically significant.





More extreme

p-value ↓

Uniformly most powerful tests

9.2.3 of Rice

07/07/2021

Uniformly most powerful tests

$$H_0 : \theta \in \Theta_0 \quad \leftrightarrow \quad H_1 : \theta \in \Theta_1$$

Definition. We say a test is of size α if $\mathrm{P}(\textit{type I error}) \leq \alpha$; that is

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} P(\mathbf{X}_n \in R \,|\, \theta) \leq \alpha.$$

Definition. A test of size α is uniformly most powerful (UMP) size α test if its power function

$$\beta(\theta) \geq \beta'(\theta) \text{ for } \theta \in \Theta_1,$$

in which $\beta'(\theta)$ is the power function of any other test of size α .

Strategy: *Maximize power* *__after__* *making sure* $\mathrm{P}(\textit{type I error}) \leq \alpha$.

Uniformly most powerful tests

Neyman-Pearson Lemma. Consider simple hypotheses $H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1.$ Among all tests that have $\mathrm{P}(\textit{type I error}) \leq \alpha$, LRT with $\mathrm{P}(\textit{type I error}) = \alpha$ has the maximum power.

⇩

LRT with significance level α *is uniformly most powerful for simple hypotheses testing*

Strategy: *Maximize power <u>after</u> making sure* $\mathrm{P}(\textit{type I error}) \leq \alpha$.

UMP one-sided tests

Example 4. Let $X_1, \ldots, X_{20}$ be i.i.d $N(\mu, 1)$. Consider testing

$$H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu = 2.$$ ← LRT rejection $R = \{\bar{X}_n \geq c'\}$.

Find the LRT with the significance level 0.05.

$\Downarrow$

$\Theta = [1, \infty)$

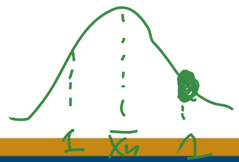
$$\to \quad H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu > 1.$$

By intuition, ~~$R = \{\bar{X}_n \geq c'\}$~~

What is LRT rejection region?

$$\lambda(\bar{X}_n) = \frac{\sup_{\Theta_0} L(\theta \mid \bar{X}_n)}{\sup_{\Theta} L(\theta \mid \bar{X}_n)} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2}}{\sup_{\mu \geq 1} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2}} = \begin{cases} \dfrac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i - \bar{X}_n)^2}}, & 1 \leq \bar{X}_n \\ 1, & 1 > \bar{X}_n \end{cases}$$

$\mu \in \mathbb{R} \to$ maximize at $\bar{X}_n$



$$\lambda(\bar{X}_n) \le c \iff e^{-\frac{1}{2}\left[\sum_{i=1}^{n}(X_i-1)^2 - \sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right]} \le c \quad \text{while} \quad \underline{1 \le \bar{X}_n}$$

$$= e^{-\frac{n}{2}(\bar{X}_n-1)^2} \le c$$

$$\Updownarrow$$

$$\frac{n}{2}(\bar{X}_n-1)^2 \ge \log 1/c$$

$$\Updownarrow$$

$$\underline{\bar{X}_n - 1} \ge \sqrt{\frac{2\log 1/c}{n}}$$

$$\Updownarrow$$

$$\bar{X}_n \ge 1 + \sqrt{\frac{2\log 1/c}{n}}.$$

UMP one-sided tests

$\mu = \mu'$, $\mu' > 1$

$$H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu = 2.$$

$\{ \bar{X}_n \geq c' \}$

LRT is UMP

has most power

$$H_0 : \mu = 1 \quad \leftrightarrow \quad H_1 : \mu > 1.$$

$\{ \bar{X}_n \geq c' \}$ has most power.

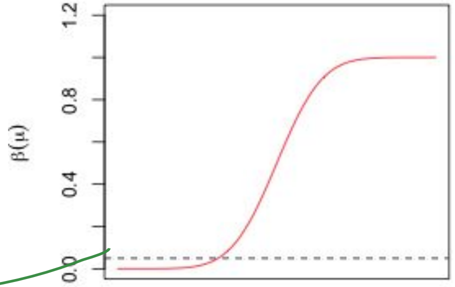
LRT is UMP

Karlin-Rubin

$$H_0 : \mu \leq 1 \quad \leftrightarrow \quad H_1 : \mu > 1.$$

LRT is uniformly most powerful!

$$\beta(\mu) = 1 - \Phi\left(\frac{c' - \mu}{1/\sqrt{n}}\right)$$



UMP one-sided tests

$$H_0 : \theta = \theta_0 \quad \leftrightarrow \quad H_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0).$$

$$\Downarrow$$

$$H_0 : \theta = \theta_0 \quad \leftrightarrow \quad H_1 : \theta > \theta_0.$$

$$\Downarrow$$ ***Karlin-Rubin***

$$H_0 : \theta \leq \theta_0 \quad \leftrightarrow \quad H_1 : \theta > \theta_0.$$

$$H_0 : \theta = \theta_0 \quad \leftrightarrow \quad H_1 : \theta = \theta_1 \quad (\theta_1 < \theta_0).$$

$$\Downarrow$$

$$H_0 : \theta = \theta_0 \quad \leftrightarrow \quad H_1 : \theta < \theta_0.$$

$$\Downarrow$$ ***Karlin-Rubin***

$$H_0 : \theta \geq \theta_0 \quad \leftrightarrow \quad H_1 : \theta < \theta_0.$$

LRT is uniformly most powerful!

Two-sided tests

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

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \quad \leftrightarrow \quad H_1 : \theta > \theta_2 \text{ or } \theta < \theta_1.$$

In general, UMP tests do not exist for two-sided hypothesis.

Tomorrow ...

- Duality of CIs and HT;
- Generalized LRT.