Lecture 13: Error Probabilities and the Power Function

Mathematical Statistics II, MATH 60062/70062

Tuesday March 8, 2022

Reference: Casella & Berger, 8.3.1-8.3.2

Recap: Hypothesis testing problem

We observe $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$, where $\theta \in \Theta$.

A **statistical hypothesis** is a statement about a population parameter θ. This statement specifies a collection of possible values of θ, i.e., the collection of distributions that $\boldsymbol{X}$ can possibly have.

In a hypothesis testing problem, two complementary hypotheses are called the **null hypothesis** (H_0) and the **alternative hypothesis** (H_1). Typically, we write

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c.$$

Recap: Hypothesis testing problem

We observe $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$, where $\theta \in \Theta$.

A **statistical hypothesis** is a statement about a population parameter θ. This statement specifies a collection of possible values of θ, i.e., the collection of distributions that $\boldsymbol{X}$ can possibly have.

In a hypothesis testing problem, two complementary hypotheses are called the **null hypothesis** (H_0) and the **alternative hypothesis** (H_1). Typically, we write

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c.$$

Usually, hypothesis tests are evaluated and compared through their probabilities of making errors.

Errors in hypothesis testing

A hypothesis test of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ might make two types of errors

- **Type I Error**: Rejecting H_0 when H_0 is true
- **Type II Error**: Not rejecting H_0 when H_1 is true.

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	☺	Type I Error
	H_1	Type II Error	☺

Power function

Let R be the rejection region for the test.

- For $\theta \in \Theta_0$,

$$P(\text{Type I Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R)$$

- For $\theta \in \Theta_0^c$,

$$P(\text{Type II Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R^c) = 1 - P_\theta(\boldsymbol{X} \in R)$$

Note that both probabilities depend on θ.

Power function

Let R be the rejection region for the test.

- For $\theta \in \Theta_0$,

This is the probability of rejecting H_0

$$P(\text{Type I Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R)$$

False Positive

- For $\theta \in \Theta_0^c$,

Note that H_0 is the prediction, the result from experiment is the truth.

$$P(\text{Type II Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R^c) = 1 - P_\theta(\boldsymbol{X} \in R)$$

True Negative

Note that both probabilities depend on θ.

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

$$\beta(\theta) = P_\theta(\boldsymbol{X} \in R).$$

Power function

Let R be the rejection region for the test.

- For $\theta \in \Theta_0$,

$$P(\text{Type I Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R)$$

- For $\theta \in \Theta_0^c$,

$$P(\text{Type II Error} \mid \theta) = P_\theta(\boldsymbol{X} \in R^c) = 1 - P_\theta(\boldsymbol{X} \in R)$$

Note that both probabilities depend on θ.

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

$$\beta(\theta) = P_\theta(\boldsymbol{X} \in R).$$

The *ideal* power function is 0 for all $\theta \in \Theta_0$ and 1 for all $\theta \in \Theta_0^c$.

Normal power function

Suppose that $X_1, \dots, X_n$ are iid $\mathcal{N}(\mu, \sigma_0^2)$, where $-\infty < \mu < \infty$ and σ_0^2 is known. Consider testing

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

The LRT rejection region is

$$R = \left\{ \boldsymbol{x} : \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}} \geq c \right\}.$$

The power function of this test is

$$\begin{aligned}
\beta(\mu) &= P_\mu \left(\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \geq c \right) \\
&= P_\mu \left(\bar{X} \geq \frac{c\sigma_0}{\sqrt{n}} + \mu_0 \right) \\
&= P_\mu \left(\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \geq \frac{\frac{c\sigma_0}{\sqrt{n}} + \mu_0 - \mu}{\sigma_0/\sqrt{n}} \right) = 1 - F_Z \left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} \right),
\end{aligned}$$

where F_Z is the standard Normal CDF.

Common practice

For testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, the Normal power function is

$$\beta(\mu) = P_\mu\left(\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \geq \frac{\frac{c\sigma_0}{\sqrt{n}} + \mu_0 - \mu}{\sigma_0/\sqrt{n}}\right) = 1 - F_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right).$$

Determine n and c such that

$$\sup_{\mu \leq \mu_0} \beta(\mu) = 0.05 \inf_{\mu \geq \mu_0 + \sigma_0} \beta(\mu) = 0.80.$$

- $P($Type I Error $\mid \mu) \leq 0.05$ for all $\mu \leq \mu_0$ (under H_0)
- $P($Type II Error $\mid \mu) \leq 0.20$ for all $\mu \geq \mu_0 + \sigma_0$ (under H_1).

First, note that $\beta(\mu)$ is an increasing function of μ since

$$\frac{\partial}{\partial\mu}\beta(\mu) = \frac{\partial}{\partial\mu}\left[1 - F_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right)\right]$$
$$= \frac{\sqrt{n}}{\sigma_0} f_Z\left(c + \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}}\right) > 0.$$

Thus,

$$\sup_{\mu \le \mu_0} \beta(\mu) = \beta(\mu_0) = 1 - F_Z(c) = 0.05 \implies c = 1.64,$$

the 95th percentile of the standard Normal. Also,

$$\inf_{\mu \ge \mu_0 + \sigma_0} \beta(\mu) = \beta(\mu_0 + \sigma_0) = 1 - F_Z(1.64 - \sqrt{n}) = 0.80$$

$$\implies 1.64 - \sqrt{n} = -0.84 \implies n = 6.15,$$

which would be rounded up to $n = 7$.

Binomial power function

Let $X \sim \mathrm{Bin}(5, \theta)$, and consider testing

$$H_0 : \theta \leq \frac{1}{2} \quad \text{versus} \quad H_1 : \theta > \frac{1}{2}.$$

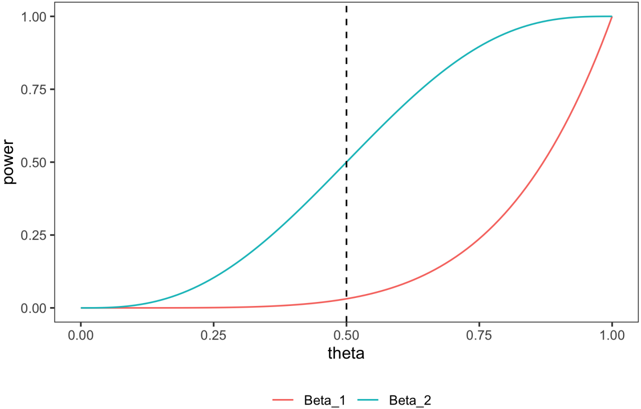
Consider two tests:

1. One rejects H_0 if and only if all "successes" are observed, and its power function is

$$\beta_1(\theta) = P_\theta(X = 5) = \theta^5.$$

2. One rejects H_0 if $X = 3$, 4, or 5, and its power function is

$$\begin{aligned}\beta_2(\theta) &= P_\theta(X \in \{3, 4, 5\}) \\ &= \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta)^1 + \binom{5}{5}\theta^5(1-\theta)^0.\end{aligned}$$


1.00
0.75
0.50
0.25
0.00
power
0.00
0.25
0.50
0.75
1.00
theta
Beta_1
Beta_2

Size and level

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **size α test** if

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

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **level α test** if

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

It is not always possible to construct a size α test (e.g., in problems that involve discrete distributions).

Size and level

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **size** α **test** if

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

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **level** α **test** if

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

It is not always possible to construct a size α test (e.g., in problems that involve discrete distributions).

Experimenters commonly specify the level of the test they wish to use, with typical choices being $\alpha = 0.01$, 0.05, and 0.1. This essentially controls the Type I Error probabilities.

Most powerful tests

Let $\mathcal{C}$ be a class of tests for testing

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta > \Theta_0^c.$$

A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a **uniformly most powerful (UMP) class $\mathcal{C}$ test** if

$$\beta(\theta) \geq \beta^*(\theta)$$

for all $\theta \in \Theta_0^c$, where $\beta^*(\theta)$ is the power function of any other test in $\mathcal{C}$.

Most powerful tests

Let $\mathcal{C}$ be a class of tests for testing

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta > \Theta_0^c.$$

A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a **uniformly most powerful (UMP) class $\mathcal{C}$ test** if

$$\beta(\theta) \geq \beta^*(\theta)$$

for all $\theta \in \Theta_0^c$, where $\beta^*(\theta)$ is the power function of any other test in $\mathcal{C}$.

We will restrict our attention to the class of all level α test. This is to avoid non-sensible tests such as one that always rejects H_0,

$$R = \{\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\}.$$

Neyman-Pearson Lemma

Consider testing

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

where H_0 and H_1 are both simple hypotheses. The PDFs/PMFs of $\boldsymbol{X} = (X_1, \ldots, X_n)$ corresponding to θ_0 and θ_1 are $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta_0)$ and $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta_1)$, respectively. Consider the test function

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_1)}{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_0)} > k \\ 0 & \text{if } \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_1)}{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_0)} < k, \end{cases}$$

for $k \geq 0$, where

$$\alpha = P_{\theta_0}(\boldsymbol{X} \in R) = E_{\theta_0}[\phi(\boldsymbol{X})].$$

Any test satisfying the above definition of $\phi(\boldsymbol{x})$ is a **(uniformly) most power level α test**.

Most powerful Binomial test

Suppose that $X \sim \mathrm{Bin}(2, \theta)$ and consider testing

$$H_0 : \theta = \frac{1}{2} \quad \text{versus} \quad H_1 : \theta = \frac{3}{4}.$$

Calculate the ratios of the Binomial PMFs:

$$\frac{f_X(0 \mid \theta = \frac{3}{4})}{f_X(0 \mid \theta = \frac{1}{2})} = \frac{1}{4}, \quad \frac{f_X(1 \mid \theta = \frac{3}{4})}{f_X(1 \mid \theta = \frac{1}{2})} = \frac{3}{4}, \quad \frac{f_X(2 \mid \theta = \frac{3}{4})}{f_X(2 \mid \theta = \frac{1}{2})} = \frac{9}{4}.$$

The Neyman-Pearson Lemma says that

- With $\frac{3}{4} < k < \frac{9}{4}$, the test that rejects H_0 if $X = 2$ is the most powerful level $\alpha = P(X = 2 \mid \theta = \frac{1}{2}) = \frac{1}{4}$ test.
- With $\frac{1}{4} < k < \frac{3}{4}$, the test that rejects H_0 if $X = 1$ or 2 is the most powerful level $\alpha = P(X = 1 \text{ or } 2 \mid \theta = \frac{1}{2}) = \frac{3}{4}$ test.