Hypothesis Testing

TONG ZHOU

JHU

1. FRAMEWORK OF TESTING

There are three steps we can employ to characterize hypothesis testing.

- **Step 1: Data and Problem.**
  1. Observe $x_1, \ldots, x_n$, and assume they are generated from a sequence of random variables $X_1, \cdots, X_n$.
- **Step 2: Statistical Modeling.**
  1. Assign $X_1, \cdots, X_n \sim P_\theta \in \mathcal{P} := \{P_\theta : \theta \in \Theta\}$. θ is assumed to be fixed and unknown, which shows the pattern of the distribution.
  2. Understand the parameter space Θ as the collection of all possible values of θ. Or more intuitively, Θ is the collection of ALL parallel worlds.
- **Step 3: Problem Formation.**
  1. Split the parameter space into Θ_0 and Θ_1, and $\Theta_0 \bigcap \Theta_1 = \emptyset$, $\Theta_0 \bigcup \Theta_1 = \Theta$.
  2. With data at hand, determine the realized world belongs to which parallel worlds.

Here are a few examples to better visualize the framework.

EXAMPLE 1.1 (Two coins example).

EXAMPLE 1.2 (Testing for ESP(Extra-Sensory Perception)).

EXAMPLE 1.3 (goodness-of-fit test (for Poisson distribution)). *This is a nonparametric testing where we must be careful to define its parameter space.*

- ***Data and problem:*** *Observe i.i.d. data $X_1, \cdots, X_n$.. Suppose that X_i 's are discrete. The problem is to test wheter observations came from a Poisson distribution.*
- ***Statistical modeling:*** $X_1, \cdots, X_n \sim P$
  The parameter space $\Theta = \{P : P \text{ is a discrete distribution}\}$.
- ***Problem formulation:***
  - H_0 : *data came from some Poisson* $\implies \Theta_0 := \{P : P \text{ is a } P(\lambda)\}$.
  - H_1 : *Data not from Poisson* $\implies \Theta_1 = \Theta \setminus \Theta_0$.

Note that both are composite hypotheses, and $\dim(\Theta_0) = 1$ *and* $\dim(\Theta_1) = \infty$.

EXAMPLE 1.4 (KS two-sample testing).
- ***Data and problem:*** *Observe i.i.d. $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$. Want to test whether they came from the same distribution.*

- ***Statistical modeling:*** *Assign* $X_1, \cdots, X_n \sim P_1$ *and* $Y_1, \cdots, Y_n \sim P_2$. *The parameter space* $\Theta = \{(P_1, P_2) : \textit{all possible distributions of } P_1 \textit{ and } P_2\}$
- ***Problem formation:*** $H_0 : P_1 = P_2$ *and* $H_1 : P_1 \neq P_2$.

2. TEST

Define $\boldsymbol{x} = (x_1, \cdots, x_n)$, and $\mathcal{X}$ is the range $(X_1, \cdots, X_n)$. Split $\mathcal{X}$ into two regions: rejection regions (denoted by $\mathcal{R}$) and accept regions. To formulate the testing procedure, we use *tesing function* $\phi : \mathcal{X} \to [0, 1]$. Often we use non-randomized test which means the range of ϕ is $\{0, 1\}$, and it is common to define $\phi(\mathcal{R}) = 1$ and $\phi(\mathcal{R}^c) = 0$.

However, we may make wrong decisions:e.g. when H_0 is true, but $\phi(\boldsymbol{x}) = 1$. We care about this error, which is called **Type I error**: $\boldsymbol{x} \in \mathcal{R}$ when Θ_0 is true. So Type I error = $\mathbb{P}(\boldsymbol{x} \in \mathcal{R}|\Theta_0) = \mathbb{P}_{\Theta_0}(\mathcal{R})$. The same for defining **Type II error**: $x \notin H_0$, when H_1 is true.

Type I and Type II error

- **Power function:**
$$\beta(\theta) := \mathbb{E}_\theta[\phi(\boldsymbol{X})] = \mathbb{P}_\theta(\mathcal{R})$$
- **Significance level:**
$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathcal{R})$$
- **Power:**
$$\beta = \beta(\theta_1) = \mathbb{E}_{\theta \in \Theta_1}[\phi(\boldsymbol{X})] = \mathbb{P}_{\theta \in \Theta_1}(\mathcal{R})$$
- **Control Type I error:** $\alpha \leqslant \bar{\alpha}$. ($\bar{\alpha}$ =0.01,0.05,0.1). (level-α test.)
- **Decision rule:** find a test to maximize the power :
$$\arg\max_\phi \beta(\theta_1) \quad s.t. \quad \alpha \leqslant \bar{\alpha}.$$

Remark 1. *Given a test, the significance level and power are* ***fixed****. Later we will see another indicator p-value, which is a random quantitly.*

Remark 2. *In general, Type I error and Type II error cannot be minimized simultaneously.*

Remark 3. *Different observations of r.v. give* <u>*different support*</u> *with varying strengths on* H_1. *So how to carve out the range of r.v is crucial.*

How to find the best test?

- **test statistics:** is a function of data, denoted by $T(X_1, \cdots, X_n)$, on which the statistical decision will be based, i.e., the rejection region $\mathcal{R}$ is determined by the values of T. So, The partitions of T's range (or rejection regions) will differ depending on different test statistics.
- **critical value:** If the rejection region is of the form $\{T > t_0\}$ or $\{T < t_0\}$, the number of t_0 is called **critical value**, which separates the rejection region and acceptance region.
- So, we summarize it

$$\text{test statistic } T \implies \text{rejection region } \mathcal{R} \implies \text{critical value } t_0 \text{ determined by } \alpha.$$

- **null distribution:** the distribution of a test statistic T under H_0.
  Why it is useful: calculate α.

p-value

The motivation is that different people may favor different tolerance level α.
p-value: the probability that a value of T that is as "extreme" or more "extreme" than the observed value of T would occur by change if the null hypothesis is true.

- **Informal definition:** $p(X) := \mathbb{P}_{H_0}(T(X) \geqslant T(x)) = \sup_{\theta\in\Theta_0} \mathbb{P}_\theta(T(X) \geqslant T(x))$
- Observe that

$$\phi_\alpha(X) = 1 \Longleftrightarrow T_{\mathsf{obs}} \in \mathcal{R} \Longleftrightarrow \text{p-value} \leqslant \alpha,$$

which motivates the **formal definition**: given $\mathcal{P}, \Theta_0$ and Θ_1, and assume if $\alpha_1 \leqslant \alpha_2 \implies \phi_{\alpha_1}(X) \leqslant \phi_{\alpha_2}(X)$ or $\mathcal{R}_{\alpha_1} \subseteq \mathcal{R}_{\alpha_2}$. Since $\phi_\alpha(X) = \mathbb{1}\{X \in \mathcal{R}_\alpha\}$, then

$$\begin{aligned} p(X) &= \inf\{\alpha : \phi_\alpha(X) = 1\} \\ &= \inf\{\alpha : T(X) \in \mathcal{R}_\alpha\} \end{aligned}$$

Here is an important property of p-value:

THEOREM 1. *If the test statistic has a continuous distribution, then under $H_0 : \theta = \theta_0$, the p-value has a uniform (0,1) distribution. Therefore, if we reject H_0 when the p-value is less than α, the probability of a type I error is α. In other words, if H_0 is true, the p-value is like a random draw from a $\mathit{Unif}(0,1)$ distribution. If H_1 is true, the distribution of the p-value will tend to concentrate closer to* 0.

PROOF. Since $T(X)$ is continuous, then for any α, the following are equivalent:

$$p(X) \leqslant \alpha \Longleftrightarrow \phi_\alpha(X) = 1 \Longleftrightarrow T(X) \geqslant t_\alpha.$$

For $\theta \in \Theta_0$,

$$\begin{aligned} \mathbb{P}_\theta(p(X) \leqslant \alpha) &= \mathbb{P}_\theta(\phi_\alpha(X) = 1) \\ &= \mathbb{P}_\theta(T(X) \geqslant t_\alpha) \\ &= \alpha. \end{aligned}$$

□

TESTING PROCEDURES

- **Step 1 (Decision):** specify the rejection region $\mathcal{R}$.
  - If $\boldsymbol{x} \in \mathcal{R}$ (or $\phi(\boldsymbol{x}) = 1$) $\implies$ Reject H_0.
  - If $\boldsymbol{x} \notin \mathcal{R}$ (or $\phi(\boldsymbol{x}) = 0$) $\implies$ Accept H_0.

3. LARGE SAMPLE TESTING

In this section, we evaluate the performance of a sequence of testings when the sample size grows. In previous sections, the sample size n is held constant, but sometimes the increase of sample size allows us to develop more testings.

Let's focus on a simple case of testing: $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$. A suitable test statistic $T_n = T(X_1, \cdots, X_n)$, and the rejection region is

$$\mathcal{R}_n = \{(x_1, \cdots, x_n) : T_n(x_1, \cdots, x_n) \geqslant C_n\},$$

where C_n is determined by

$$\mathbb{P}_{\theta_0}(T_n \geqslant C_n) = \alpha.$$

But usually this condition is replaced by a weaker condition and an asymptotic level α satisfy

$$\mathbb{P}_{\theta_0}(T_n \geqslant C_n) \longrightarrow \alpha \quad \text{as} \quad n \longrightarrow \infty.$$

Here α is the pre-assigned level of significance which controls the probability of falsely rejecting the null hypothesis when in fact it is true.

Definition 1 (consistent). *The sequence of tests defined above is said to be consistent against the alternative θ if*

$$\beta_n(\theta) \longrightarrow 1 \quad \text{as} \quad n \longrightarrow \infty,$$

where $\beta_n(\theta)$ is the power *of the test T_n.*

¡++¿

4. EXERCISES

Exercise 1. *Testing the value of the parameter p of a Bernoulli distribution with 10 trials. Let X be the number of success, then $X \sim B(10, p)$. Suppose the rejection region $\mathcal{R} = \{7, 8, 9, 10\}$.*

1. *For the following cases, calculate their type I error and type II error:*
   - *If $H_0 : p = 0.5$ VS $H_1 : p = 0.7$.*
   - *If $H_0 : p = 0.5$ VS $H_1 : p > 0.5$.*
   - *If $H_0 : p \leqslant 0.5$ VS $H_1 : p > 0.5$.*
2. *For the third case, draw a graph to show the relationship between the power function $\beta(p)$ and the true p. Discuss how the calculated type I and type II errors correspond to regions and segments in the drawn graph. Also discuss why H_0 is more protected.*

Exercise 2. *Let $X_1, \cdots, X_n \sim \mathsf{Uniform}(0, \theta)$ and let $Y = \max\{X_1, \cdots, X_n\}$. We want to test*

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

The Wald test is not appropriate since Y does not converge to a Normal. Suppose we decide to test this hypothesis by rejecting H_0 when $Y > c$.

(i) Find the power function.
(ii) What choice of c will make the size of the test 0.05.
(iii) In a sample of size $n = 20$ with $Y = 0.48$ what is the p-value? What conclusion about H_0 would you make?
(iv) In a sample of size $n = 20$ with $Y = 0.52$ what is the p-value? What conclusion about H_0 would you make?

PROOF. (i) By definition, the power function is

$$\begin{aligned}\beta(\theta) &= \mathbb{P}_\theta(\mathcal{R})\\ &= \mathbb{P}_\theta(Y > c)\\ &= 1 - \mathbb{P}_\theta(Y \leqslant c)\\ &= 1 - \mathbb{P}_\theta\left(\max\{X_1, \cdots, X_n\} \leqslant c\right)\\ &= 1 - \prod_{i=1}^{n} \mathbb{P}_\theta(X_i \leqslant c)\end{aligned}$$

(ii) The Type I error by definition is

$$\alpha = \beta(\theta = 1/2) = 1 - c^n$$

Let $\alpha = 0.05$, then c can be recovered by

$$c = (1 - .05)^{1/n}$$

□

EXERCISE 3. *Let $X_1, \cdots, X_n \sim \mathcal{N}(\theta, 1)$. Consider testing*

$$H_0 : \theta = 0 \textit{ versus } H_1 : \theta = 1$$

Let the rejection region be $\mathcal{R} = \{x^n : T(x^n) > c\}$ where $T(x^n) = n^{-1}\sum_{i=1}^{n} X_i$.

(i) Find c so that the test has size α.
(ii) Find the power under H_1, that is, find $\beta(1)$.
(iii) Show that $\beta(1) \to 1$ as $n \to \infty$.

PROOF. (i) The size is

$$\alpha = \mathbb{P}_{\theta_0}(\mathcal{R}) = \mathbb{P}\left(n^{-1}\sum_{i=1}^{n} X_i > c | \theta = 0\right) = 1 - \Phi(nc).$$

So c can be recovered by

$$c = \frac{\Phi^{-1}(1-\alpha)}{n}.$$

(ii) By definition

$$\begin{aligned}\beta(1) &= \mathbb{P}_{\theta_1}(\mathcal{R}) \\ &= \mathbb{P}_{\theta_1}\left(n^{-1}\sum_{i=1}^{n} X_i > c\right) \\ &= 1 - \Phi(n(c-1))\end{aligned}$$

(iii) Observe that as $n \to \infty$, $c \to 0$, so that $\Phi(n(c-1)) \to 0$. It follows that

$$\beta(1) \to 1.$$

□

Exercise 4. *Let $X_1, \cdots, X_n$ be i.i.d. with density f which can be either f_0 or else f_1, where f_0 is Poisson $P(1)$ and f_1 is the Geometric distribution with $p = \frac{1}{2}$. Find the most powerful (MP) test of the hypothesis $H_0 : f = f_0$ versus $H_1 : f = f_1$ at level of significance $\alpha = .05$. [Hint: apply Neyman-Pearson Lemma and construct a randomized test.]*

Exercise 5. *Let $X_1, \cdots, X_n$ be independent random variables with density f given by*

$$f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad \text{for } x \geqslant 0,$$

where $\theta \in \Omega = (0, \infty)$. Derive the uniformly most powerful (UMP) test for testing the hypothesis $H_0 : \theta \geqslant \theta_0$ versus $H_1 : \theta < \theta_0$ at level of significance α. [Hint: $\sum_{i=1}^{n} X_i$ follows a Gamma distribution.]

Exercise 6. *Let X_1, X_2, X_3 be i.i.d. from Binomial $B(1, p)$. Derive the uniformly most powerful unbiased (UMPU) test for testing the hypothesis $H_0 : p = .25$ versus $H_1 : p \neq .25$ at level of significance α. Determine the test for $\alpha = .05$. [Hint: if $Y \sim B(3, 0.25)$, then $\mathbb{P}(Y = 0) = 0.42$, $\mathbb{P}(Y = 1) = 0.42$, $\mathbb{P}(Y = 2) = 0.14$, and $\mathbb{P}(Y = 3) = 0.02$.]*

Exercise 7. *Let $X_1, \cdots, X_n$ be a random sample from an exponential distribution with the density $f(x|\theta) = \theta e^{-\theta x}$. Derive a likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, and show that the rejection region of the form $\left\{\bar{X}e^{-\theta_0\bar{X}} \leqslant c\right\}$.*

¡++¿