Hypothesis Testing

Andrew Nobel

October, 2024

Formulation of the Problem

Hypothesis Testing

Ingredients

- Family $\mathcal{P} = \{f(\cdot|\theta) : \theta \in \Theta\}$ of distributions on sample space $\mathcal{X}$
- Distinguished subset $\Theta_0 \subseteq \Theta$. Let $\Theta_1 = \Theta_0^c$

Task: Use observation $X \sim f(\cdot|\theta) \in \mathcal{P}$ to distinguish between hypotheses

- $\mathrm{H}_0 : \theta \in \Theta_0$ Null hypothesis
- $\mathrm{H}_1 : \theta \in \Theta_1$ Alternative hypothesis

Asymmetry: The null hypothesis represents default or status quo. We reject H_0 (accept H_1) only when there is significant evidence to do so.

- (Medicine) H_0: New drug/procedure no better than existing drug/procedure
- (Quality control) H_0: Proportion of defectives acceptable
- (Criminal trial) H_0: Defendant is innocent

Hypothesis Testing, cong

Terminology: An element $\theta \in \Theta_1$ is referred to as an *alternative*

Definition: H_i is *simple* if $|\Theta_i| = 1$, *composite* otherwise. For $\Theta \subseteq \mathbb{R}$ hypothesis H_i is

- one-sided if $\Theta_i = \{\theta \in \Theta : \theta \leq \theta_a\}$ or $\Theta_i = \{\theta \in \Theta : \theta \geq \theta_b\}$
- two-sided if $\Theta_i = \{\theta \in \Theta : \theta \leq \theta_a \text{ or } \theta \geq \theta_b\}$

Example: $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta, 1)$ with $\theta \in \Theta \subseteq \mathbb{R}$

(a) $\Theta = \{0, 1\}$. Test $\text{H}_0 : \theta = 0$ vs. $\text{H}_1 : \theta = 1$

(b) $\Theta = [0, \infty)$. Test $\text{H}_0 : \theta = 0$ vs. $\text{H}_1 : \theta > 0$

(c) $\Theta = \mathbb{R}$ with θ_0 fixed. Possible tests

$$\text{H}_0 : \theta \leq \theta_0 \ \text{ vs. } \ \text{H}_1 : \theta > \theta_0$$

$$\text{H}_0 : \theta = \theta_0 \ \text{ vs. } \ \text{H}_1 : \theta \neq \theta_0$$

Hypothesis Tests

Hypothesis test: Specified by a pair (T, R)

- Test statistic $T : \mathcal{X} \to \mathcal{Y}$
- Rejection region $R \subseteq \mathcal{Y}$

Test rejects H_0 iff $T(x) \in R$. Associated decision rule $\phi : \mathcal{X} \to \{0, 1\}$ given by

$$\phi(x) \; = \; \mathbb{I}(T(x) \in R) \; = \; \begin{cases} 1 & \text{accept } H_1 \\ 0 & \text{accept } H_0 \end{cases}$$

Note

- Test partitions $\mathcal{X}$ into $\mathcal{X}_1 = \{x : T(x) \in R\}$ and $\mathcal{X}_0 = \{x : T(x) \notin R\}$
- Rejection region R will depend on H_0 (e.g., one-sided vs two-sided) and desire to control probability of falsely rejecting H_0

Likelihood Ratio Tests

Testing Simple vs Simple Hypotheses

Setting: Suppose $\Theta = \{\theta_0, \theta_1\}$. Let $f_0 = f(\cdot|\theta_0)$ and $f_1 = f(\cdot|\theta_1)$

- $\mathrm{H}_0 : \theta = \theta_0$, equivalently, observation $X \sim f_0$
- $\mathrm{H}_1 : \theta = \theta_1$, equivalently, observation $X \sim f_1$

Formulating a Test

- Natural test statistic is the likelihood ratio of f_1 vs f_0

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)} = \frac{L(\theta_1|x)}{L(\theta_0|x)}$$

- Natural rejection region is $R = [\tau, \infty)$, as larger values of $\Lambda(x)$ favor H_1
- Test (Λ, R) rejects H_0 if ratio $f_1(x)/f_0(x) \geq \tau$

General Likelihood Ratio Tests

Idea: Compare max of $L(\theta|x)$ over Θ_0 with max over full parameter space Θ

Definition: The *likelihood ratio* test statistic is

$$\lambda(x) = \frac{\max_{\theta \in \Theta_0} L(\theta|x)}{\max_{\theta \in \Theta} L(\theta|x)} \ \in \ [0, 1]$$

The *likelihood ratio test* (LRT) rejects H_0 if $\lambda(x) \leq c$ for fixed $c \in [0, 1]$

Note

- Large values of $\lambda(x)$ favor H_0, small values favor H_1
- $\lambda(x) = 1$ iff $\max_{\theta \in \Theta_0} L(\theta|x) \geq \max_{\theta \in \Theta_1} L(\theta|x)$
- $\lambda(x) = L(\hat{\theta}_0|x)/L(\hat{\theta}|x)$ where $\hat{\theta}_0 =$ MLE for Θ_0 and $\hat{\theta} =$ MLE for Θ

LRT for Normal Mean, Known Variance

Let $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$ with $\theta \in \mathbb{R}$ and σ^2 known. Consider testing

$$\mathrm{H}_0 : \theta = \theta_0 \ \text{ vs. } \ \mathrm{H}_1 : \theta \neq \theta_0$$

Fact. LRT is equivalent to a two-sided z-test: $\lambda(x) \leq c$ if and only if

$$T(x) \ = \ \frac{|\overline{x} - \theta_0|}{\sigma/\sqrt{n}} \ \geq \ \tau(c)$$

for some number $\tau(c) \geq 0$

LRT for Normal Mean, Unknown Variance

Given $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ unknown. Consider testing

$$\mathrm{H}_0 : \mu \leq \mu_0 \ \text{ vs. } \ \mathrm{H}_1 : \mu > \mu_0$$

Definition: In this problem variance σ^2 is said to be a *nuisance parameter*. It is unknown but is not the quantity we are trying to make inferences about

Fact. LRT is equivalent to a one-sided t-test: $\lambda(x) \leq c$ iff

$$T(x) \ = \ \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \ \geq \ \tau(c)$$

for some number $\tau(c) \geq 0$

LRT and Sufficient Statistics

Let $T : \mathcal{X} \to \mathcal{Y}$ be sufficient for θ, and let $T(X) \sim g(y|\theta)$ when $X \sim f(x|\theta)$

- Let $\tilde{L}(\theta|y) := g(y|\theta)$ be the likelihood function of θ when $T(x) = y$
- Define likelihood ratio test statistic based on the likelihood $\tilde{L}(\theta|y)$

$$\tilde{\lambda}(y) = \frac{\max_{\theta \in \Theta_0} \tilde{L}(\theta|y)}{\max_{\theta \in \Theta} \tilde{L}(\theta|y)}$$

The factorization theorem shows that $\lambda(x) = \tilde{\lambda}(T(x))$ for each $x \in \mathcal{X}$

Upshot: We can always construct a LRT based on a sufficient statistic

Evaluating Test Performance: Level and Power

Evaluating Test Performance

Recall: Test $\mathrm{H}_0 : \theta \in \Theta_0$ vs. $\mathrm{H}_1 : \theta \in \Theta_1$ described by decision rule $\phi : \mathcal{X} \to \{0, 1\}$

- $\phi(x) = 0$ means we accept H_0 (equivalently, reject H_1)
- $\phi(x) = 1$ means we reject H_0 (equivalently, accept H_1)

Under the 0/1 loss a rule ϕ can make two kinds of errors (each with loss 1)

- Type I: Reject H_0 when H_0 true (i.e. $\phi(x) = 1$ when $\theta \in \Theta_0$)
- Type II: Accept H_0 when H_1 true (i.e. $\phi(x) = 0$ when $\theta \in \Theta_1$)

Truth / Decision	H_0	H_1
H_0	Correct	Type I
H_1	Type II	Correct

Power Function of a Test

Definition: The *power function* $\beta : \Theta \to [0,1]$ of a test (T, R) is

$$\beta(\theta) \;=\; \mathbb{P}_\theta(\text{Test rejects H}_0) \;=\; \mathbb{P}_\theta(T(X) \in R)$$

When the test is specified by a rule $\phi : \mathcal{X} \to \{0,1\}$ we can write

$$\beta(\theta) \;=\; \mathbb{P}_\theta(\phi(X) = 1) \;=\; \mathbb{E}_\theta \phi(X)$$

Power function contains all relevant information about error probabilities of the test

- If $\theta \in \Theta_0$ then $\beta(\theta) = P_\theta(\text{Type I error})$
- If $\theta \in \Theta_1$ then $\beta(\theta) = 1 - P_\theta(\text{Type II error})$, *power* of test at alternative θ

Competing Goals: We would like

- $\beta(\theta)$ close to zero for $\theta \in \Theta_0$ (small Type I error)
- $\beta(\theta)$ close to one for $\theta \in \Theta_1$ (large power against alternatives)

Level, Size, Unbiasedness

Standard approach in frequentist hypothesis testing is to maximize power of a test subject to a bound on its Type I error

Definition: Let $\phi : \mathcal{X} \to \{0, 1\}$ be a rule having power function $\beta(\theta)$ and let $\alpha \in [0, 1]$

- ϕ is a *level*-α test if $\max_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$
- ϕ is a *size*-α test if $\max_{\theta \in \Theta_0} \beta(\theta) = \alpha$
- ϕ is *unbiased* if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \inf_{\theta \in \Theta_1} \beta(\theta)$

Examples

Ex 1. (Binomial mean) Let $X \sim \text{Bin}(n, \theta)$ with $\theta \in [0, 1]$. Consider testing

$$\text{H}_0 : \theta \leq 1/2 \;\text{ vs. }\; \text{H}_1 : \theta > 1/2$$

Smaller values of X favor H_0 so consider rejection regions $R_k = \{k, k+1, \ldots, n\}$.

Of interest: Power function $\beta_k(\theta)$ of test (X, R_k).

Ex 2. (Normal mean) Let $X_1, \ldots, X_n$ be i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$ with σ^2 known. Consider

$$\text{H}_0 : \theta \leq \theta_0 \;\text{ vs. }\; \text{H}_1 : \theta > \theta_0$$

In this case LRT is equivalent to (T, R_c) where

$$T(x) \;=\; \frac{\overline{x} - \theta_0}{\sigma/\sqrt{n}} \qquad\qquad R_c = [c, \infty)$$

Of interest: Power function $\beta_c(\theta)$ of test (T, R_c).

Uniformly Most Powerful Tests

Neyman-Pearson Lemma

Uniformly Most Powerful Tests

Definition: Test ϕ is a uniformly most powerful (UMP) level α test for H_0 vs. H_1 if

(1) The test ϕ has level α

(2) The test ϕ has greater power than any other level α test. If ϕ' has level α then

$$\beta(\theta) := \mathbb{E}_\theta \phi(X) \geq \mathbb{E}_\theta \phi'(X) := \beta'(\theta)$$

for all alternatives $\theta \in \Theta_1$

Neyman-Pearson Theorem

Setting: Family $\mathcal{P} = \{f_0, f_1\}$ consisting of two densities on $\mathbb{R}$

- Given observation $X \sim f_\theta$ wish to test $\mathrm{H}_0 : \theta = 0$ vs $\mathrm{H}_1 : \theta = 1$
- In this setting, for any test $\phi : \mathcal{X} \to \{0, 1\}$ we have

$$\text{size}(\phi) = \mathbb{E}_0 \phi(X) \ \text{ and } \ \text{power}(\phi) = \mathbb{E}_1 \phi(X)$$

- Natural test statistic is the likelihood ratio

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)} = \frac{\text{likelihood of } \theta = 1 \text{ given } x}{\text{likelihood of } \theta = 0 \text{ given } x}$$

- Natural test ϕ^* rejects H_0 when $\Lambda(x) > \tau$, accepts H_0 when $\Lambda(x) < \tau$

Neyman Pearson Theorem

Theorem: Let $\phi^* : \mathcal{X} \to \{0, 1\}$ be a test for $\mathrm{H}_0 : \theta = 0$ vs $\mathrm{H}_1 : \theta = 1$ such that

$$\phi^*(x) = 1 \text{ if } \Lambda(x) > \tau \qquad \phi^*(x) = 0 \text{ if } \Lambda(x) < \tau$$

Suppose that ϕ^* has size α and let $\phi : \mathcal{X} \to \{0, 1\}$ be any other test.

Optimality: If size$(\phi) \leq$ size$(\phi^*) = \alpha$, then power$(\phi) \leq$ power(ϕ^*)

Uniqueness: If ϕ has the same level and power as ϕ^* then $\phi(x) = \phi^*(x)$ outside the set of x such that $\Lambda(x) = \tau$

Note

- Conclusion of Optimality is that ϕ^* is UMP level α test for H_0 vs H_1
- Size α of ϕ^* depends on τ and value of ϕ^* when $\Lambda(x) = \tau$
- If we allow randomization, we can find a UMP test of any size $\alpha \in (0, 1)$

P-Values

Standard Definition

Setting: Testing $\mathrm{H}_0 : \theta \in \Theta_0$ vs. $\mathrm{H}_1 : \theta \in \Theta_1$ using statistic $S : \mathcal{X} \to \mathbb{R}$ with larger values favoring the alternative

- Observation $X \sim P_\theta$ yields data $x \in \mathcal{X}$ and test statistic $S(x)$
- Task: Quantify the strength of evidence against H_0 provided by $S(x)$
- Consider largest (worst case) probability that $S(X)$ exceeds $S(x)$ under the null H_0. Result is p-value

$$p(x) \;=\; \max_{\theta \in \Theta_0} \mathbb{P}_\theta(S(X) \geq S(x))$$

Fact: For each $\theta \in \Theta_0$ and $\alpha \in (0,1)$ we have $\mathbb{P}_\theta(p(X) \leq \alpha) \leq \alpha$, equivalently, $p(X)$ is stochastically larger than a Unif$(0,1)$ random variable