Hypothesis Testing: Part 1

Dr. Qiuzhuang Sun

STAT3023

General setup and definitions

- Examples of statistical hypothesis testing problems:
  - decide based on clinical trial data (control vs. treatment) whether a new drug lowers blood pressure
  - decide if the lifetime of a mechanical component in a car follows an exponential distribution or another distribution
- Treat observed data as values taken by random variables (often assumed to be iid)
- Statistical hypothesis: an assertion or conjecture about the underlying **distribution** of the random variables
- Null hypothesis H_0 and alternative hypothesis H_1: a distribution (or a family of distributions) to be compared

General setup and definitions

- Test statistic: a real-valued function $T(\mathbf{X})$ of the data $\mathbf{X} = (X_1, \ldots, X_n)$, capturing certain features of the distribution.
- p-value: assuming H_0 true,

$$P(\text{at least as much evidence against } H_0 \text{ as was observed}),$$

  (or $\sup_{P \in H_0} P(\cdots)$). The smaller the p-value, the stronger the evidence against H_0.

General setup and definitions

- Goal: **compare different tests**
- A test procedure partitions possible values of $\mathbf{X} = (X_1, \ldots, X_n)$ into two subsets: an acceptance region and a **rejection region (or critical region)**, denoted by C, assuming H_0:

  Reject H_0 if $\mathbf{X} \in C$
  Accept H_0 if $\mathbf{X} \in C^c$

Two types of errors

- Type I and Type II errors

	H_0 is true	H_0 is false
Accept H_0	No error	Type II error: P_{H_1}(accept H_0)
Reject H_0	Type I error: P_{H_0}(reject H_0)	No error

- power $= 1 -$ Type II error $= P_{H_1}$(reject H_0)
- Tradeoff between Type I and Type II errors

Optimality of tests

- A level-α test is any test such that

$$P_{H_0}(\text{reject } H_0) \leq \alpha.$$

- The mathematical framework developed by Neyman and Pearson allows us to identify **optimal** level-α tests in certain scenarios, meaning that the power of the test is as high as possible.

Simple and composite hypothesis

- Usually the hypothesis H_0 is a subset of a larger statistical model $\mathcal{M}$.
- The complement of H_0 within $\mathcal{M}$ is called the alternative hypothesis. i.e., $\mathcal{M} = H_0 \cup H_1$, and $H_0 \cap H_1 = \emptyset$.
- A hypothesis containing only one distribution is called simple; if it contains more than one distribution it is called composite.
- We will consider 3 cases:
    - Simple H_0 vs simple H_1: optimal tests can be easily found
    - Simple H_0 vs composite H_1
    - Composite H_0 vs simple H_1

  The optimal tests exist only in special cases for the latter two.

Simple vs Simple: the NP Lemma

Likelihood ratio statistic and critical region:

Consider $H_0 : \mathbf{X} \sim f_0(\cdot)$ and $H_1 : \mathbf{X} \sim f_1(\cdot)$

- $f_0(\cdot)$, $f_1(\cdot)$ are PDFs for continuous distributions or PMFs for discrete distributions

The likelihood ratio statistic is

$$Y = \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})}.$$

The larger this ratio, the more likely f_1 is the underlying distribution.

The critical region has the form $C = \left\{\mathbf{x} : \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} \geq y\right\}$ for some y.

Simple vs Simple: the NP Lemma

Recall

$$Y = \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})} \quad \text{and} \quad C = \left\{ \mathbf{x} : \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} \geq y \right\}.$$

The Neyman-Pearson (NP) Lemma: Let H_0 and H_1 be simple hypotheses (in which the distributions are either both discrete or both continuous). Fix a level $0 < \alpha < 1$, and suppose there exists y_α such that

$$P_{f_0}(\mathbf{X} \in C) = P_{f_0}(Y \geq y_\alpha) = \alpha.$$

Then for any other test of H_0 with significance level at most α, its power against H_1 is at most the power of this likelihood ratio test.

Simple vs Simple: the NP Lemma

Proof:

Simple vs Simple: the NP Lemma

Simple vs Simple: the NP Lemma

Example 1

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0, 1)$ and $H_1 : X_i \sim N(\mu_0, 1)$ for some fixed $\mu_0 \neq 0$. What is the most powerful likelihood ratio test?

Example 1

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(\mu_0, 1)$ for some fixed $\mu_0 \neq 0$. What is the most powerful likelihood ratio test?

Example 1

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(\mu_0, 1)$ for some fixed $\mu_0 \neq 0$. What is the most powerful likelihood ratio test?

Example 1

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(\mu_0, 1)$ for some fixed $\mu_0 \neq 0$. What is the most powerful likelihood ratio test?

Example 2

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(0, \sigma_0^2)$ for some known $\sigma_0^2 \neq 1$. What is the most powerful likelihood ratio test?

Example 2

Let $X_1, \dots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(0, \sigma_0^2)$ for some known $\sigma_0^2 \neq 1$. What is the most powerful likelihood ratio test?

Example 2

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(0, \sigma_0^2)$ for some known $\sigma_0^2 \neq 1$. What is the most powerful likelihood ratio test?

Example 3

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(\mu_0, \sigma_0^2)$ for some known $\mu_0 \neq 0$ and $\sigma_0^2 \neq 1$.

Example 3

Let $X_1, \ldots, X_n$ be iid RVs. Consider $H_0 : X_i \sim N(0,1)$ and $H_1 : X_i \sim N(\mu_0, \sigma_0^2)$ for some known $\mu_0 \neq 0$ and $\sigma_0^2 \neq 1$.

The discrete case: randomised tests

If $Y = \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})}$ has a discrete distribution, there may be no exact value y_α such that $P_0(Y \geq y_\alpha) = \alpha$ for **any** given $0 < \alpha < 1$. We can get around this in a couple of ways:

1. restrict attention to only those values of α such that the above holds for some y_α;
2. introduce the concept of a randomised test.

Example 4

Suppose $X \sim \text{Bin}(5, p)$. Consider testing $H_0 : p = 1/2$ vs. $H_1 : p = 3/4$.

Test functions

Let $\delta(\cdot)$ be a function of $\mathbf{X}$ taking values in $[0,1]$. If we observe the value $\mathbf{x}$, we reject H_0 with probability $\delta(\mathbf{x})$. $\delta(\cdot)$ is called a test function. Formally let $U \sim U[0,1]$, independent of $\mathbf{X}$. Then we reject if $U \leq \delta(\mathbf{X})$.

Example 4 (continued)

Randomised tests

In general, writing $f_0(\mathbf{x}) = P_0(\mathbf{X} = \mathbf{x})$ and $f_1(\mathbf{x}) = P_1(\mathbf{X} = \mathbf{x})$, the most powerful level-α test for $H_0 : \mathbf{X} \sim P_0(\cdot)$ vs. $H_1 : \mathbf{X} \sim P_1(\cdot)$ is given by

$$\delta(\mathbf{x}) = \begin{cases} 1, & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > y, \\ \gamma, & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = y, \\ 0, & \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} < y, \end{cases}$$

where γ, y are chosen such that $E_0[\delta(\mathbf{X})] = \alpha$.

Specifically, y satisfies $P_0(Y \geq y) \geq \alpha > P_0(Y > y)$ and γ satisfies

$$E_0[\delta(\mathbf{X})] = \gamma P_0(Y = y) + P_0(Y > y) = \alpha,$$

yielding $\gamma = \frac{\alpha - P_0(Y > y)}{P_0(Y = y)}$.

In the continuous case: $P_0(\text{reject } H_0) = P_0(\delta(\mathbf{X}) = 1)$.

Example 5

Consider $H_0 : X \sim \text{Bin}(5, \frac{1}{2})$ vs. $H_1 : X \sim \text{Bin}(5, p_0)$ for some $p_0 \neq \frac{1}{2}$.

Example 5

Consider $H_0 : X \sim \text{Bin}(5, \frac{1}{2})$ vs. $H_1 : X \sim \text{Bin}(5, p_0)$ for some $p_0 \neq \frac{1}{2}$.

Simple vs. composite: UMP tests

Now consider testing simple H_0 against composite H_1. Suppose we have a parametric family of distributions depending on 1 parameter θ:

$$\{f_\theta(\cdot) : \theta \in \Theta\}$$

for some $\Theta \subseteq \mathbb{R}$ and we model data $\mathbf{x}$ as observed values of $\mathbf{X} \sim f_\theta(\cdot)$, θ is unknown. We wish to test $H_0 : \theta = \theta_0$ for some $\theta_0 \in \Theta$ vs. $H_1 : \theta \in \Theta \backslash \{\theta_0\}$. We would like to find a uniformly most powerful (UMP) test (at some level α) $\delta_0(\cdot)$ such that

- $E_{\theta_0}[\delta_0(\mathbf{X})] = \alpha$ (level-α test)
- $E_{\theta}[\delta_0(\mathbf{X})] \geq E_{\theta}[\delta_1(\mathbf{X})]$ for all other test function with $E_{\theta_0}[\delta_1(\mathbf{X})] \leq \alpha$ and $\theta \in \Theta \backslash \{\theta_0\}$.

Simple vs. composite: UMP tests

- Examples 1, 2, 5 all had a special property which we now formalise
- A parametric family $\{f_\theta(\cdot) : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}$, is said to have monotone likelihood ratio in a statistic $T(\cdot)$ if for any $\theta_0 < \theta_1$, both in Θ:
    - $f_{\theta_0}, f_{\theta_1}$ are different distributions
    - $\frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}$ is a nondecreasing function of $T(\mathbf{x})$

Check for Example 1: $\{\mathbf{X} = (X_1, \ldots, X_n),\ X_i \sim N(\mu, 1),\ \mu \in \mathbb{R}\}$

Exercise: check the parametric families in Examples 2 and 5 have a monotone likelihood ratio in a statistic $T(\cdot)$ that you identified

Simple vs. composite: UMP tests

Theorem (one-sided tests): Suppose a family $\{f_\theta(\cdot) : \theta \in \Theta\}$ has monotone likelihood ratio in the statistic $T(\mathbf{X})$. Then for any $\theta_0 \in \Theta$, a UMP level-α test exists for testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$ and is given by

$$\delta(\mathbf{X}) = \begin{cases} 1, & T(\mathbf{X}) > c, \\ \gamma, & T(\mathbf{X}) = c, \\ 0, & T(\mathbf{X}) < c, \end{cases}$$

where c, γ are chosen such that $E_0[\delta(\mathbf{X})] = \alpha$.

How about the case that the likelihood ratio is nonincreasing in $T(\mathbf{X})$? The UMP test exists for $H_1 : \theta < \theta_0$.

Simple vs. composite: UMP tests

Proof:

Simple vs. composite: UMP tests

Example: Consider a 1-parameter exponential family with $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$ and PDF

$$f_\theta(x) = h(x)\exp(\eta(\theta)T(x) - A(\theta)).$$

If $\eta(\theta)$ is strictly increasing in θ, then the likelihood ratio is increasing in $T(x)$.

Simple vs. composite: UMP tests

Example: Suppose X is the number of faulty items in a random sample of size n taken without replacement from a batch of size N. Let M be the total no. of faulty items in the batch. Consider testing $H_0 : M = M_0$ vs. $M > M_0$. Find the UMP test.

Here, X has a hypergeometric distribution with

$$P_M(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

Simple vs. composite: UMP tests

Example: Suppose X is the number of faulty items in a random sample of size n taken without replacement from a batch of size N. Let $M=$ total no. of faulty items in the batch. Consider testing $H_0 : M = M_0$ vs $M > M_0$. Find the UMP test.

Simple vs. composite: UMP tests

Two-sided tests

Now consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. It can be shown (proof omitted) that if the family has monotone likelihood ratio then the power functions of the UMP 1-sided tests are strictly monotone.

Each of the 1-sided UMP test works well for their stated 1-sided H_1. However, for the 2-sided H_1, there exists θ such that the tests are biased in the sense that

$$E_\theta[\delta_0(\mathbf{X})] < E_{\theta_0}[\delta_0(\mathbf{X})].$$