

Hypothesis Testing

Def: A hypothesis is a statement about a population parameter.

Two complementary hypotheses in one H-T problem are called null hypothesis H_0 and alternative hypothesis H_A (or H_1)

$$\text{e.g. } A \subseteq \Omega, H_0 = \theta \in A, H_1 = \theta \in A^c.$$

(1) Method of Finding Tests:

The procedure of hypothesis testing is:

After observing the samples taken, make a decision on accepting H_0 or H_1 . The subset of " H_0 is rejected" of sample space is rejection region R .

Remark: If $H_0: \theta = 1.5$, $H_1: \theta > 0.5$. We will not simply

calculate the estimate of θ , i.e. $\hat{\theta}$. Compare $\hat{\theta}$ → That's also why we put $\theta = 0.0$ on H_0 .
with θ . Since if $\hat{\theta} = 0.500001$, it's vague to say $\hat{\theta}$ falls into H_0 or H_1 .
A little deviation can be tolerated

Moreover, H_0 and H_1 are asymmetric. We're pretending to protect H_0 .

① Likelihood Ratio Test

Def: Likelihood Ratio Test statistic for $H_0: \theta \in \Theta_0$.

vs. $H_1: \theta \in \Theta_0^c$ is

$$\lambda(\vec{x}) = \frac{\sup_{\Theta_0} L(\theta | \vec{x})}{\sup_{\Theta_0^c} L(\theta | \vec{x})}$$

Likelihood Ratio Test is any test has rejection region of form $\{\vec{x} | \lambda(\vec{x}) < c\}$, $0 < c < 1$.

A computational simplification:

Thm. $T(\vec{x})$ is sufficient statistic for θ . $T(\vec{x}) \sim g(t|\theta)$

$$\lambda^*(t) = \frac{\sup_{\Theta_0} g(t|\theta)}{\sup_{\Theta_0^c} g(t|\theta)}. \text{ Then } \lambda(\vec{x}) = \lambda^*(T(\vec{x})), \forall \vec{x} \in \mathcal{X}.$$

Pf: By Factorization Thm.

③ Bayesian Tests:

Given $\theta \sim \pi(\cdot)$, $X \sim f(x|\theta)$. Calculate $Z(\theta | \vec{x})$

Then we have: $P(\theta \in \Theta_0 | \vec{x})$, $P(\theta \in \Theta_0^c | \vec{x})$

since these two prob have sum 1.

We can set rejection region: $R = \{\vec{x} | P(\theta \in \Theta_0^c | \vec{x}) > c\}$
for some const. $c \in (0, 1)$

③ Union-Intersection Test and

Intersection-Union Test:

- Test for complicated null hypotheses can be

developed from simpler null hypotheses.

i) H-I Method:

For $H_0: \theta \in \bigcap_{y \in I} \mathcal{D}_y$, we can separate it:

Test: $H_0: \theta \in \mathcal{D}_y$, v.s. $H_1: \theta \in \mathcal{D}_y^c$, with
rejection region R_y

So, the rejection region of H_0 is $R = \bigcup_{y \in I} R_y$

ii) I-H Method:

Analogously, $H_0: \theta \in \bigcup_{y \in I} \mathcal{D}_y$, separate:

$H_0: \theta \in \mathcal{D}_y$ v.s. $H_1: \theta \in \mathcal{D}_y^c$, with
rejection region R_y .

Then $R = \bigcap_{y \in I} R_y$

(2) Methods of evaluating tests:

Intuitively, hypothesis test are evaluated by comparing the prob. of making mistake.

① Error prob and

Power function:

		Decision	
		Accept H_0	Reject H_0
		True H_0	True H_1
	H_0	Correct Decision	Type I error
	H_1	Type II error	Correct Decision

$$\therefore \begin{cases} P(\text{Type I error}) = P(X \in R | \theta \in \Theta_0) \\ P(\text{Type II error}) = P(X \notin R^c | \theta \in \Theta_0^c) \end{cases}$$

Def: The power function of hypothesis test with rejection Region R is: $\beta(\theta) = P_{\theta}(X \in R)$

\Rightarrow For $0 < \alpha < 1$, a test with power function

$\beta(\theta)$ is "size" α test. If $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.

is a "level" α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.

Remark: It's usually impossible to make both error probs small. We limit on the class with small $P(\text{Type I})$ firstly, then minimize $P(\text{Type II})$

\Rightarrow It's important to specify null and alternative hypotheses for controlling $P(\text{Type I error})$

- Methods of determining the null hypothesis H_0
 - i) Choose simpler hypo (e.g. $\theta = 0.5$) in H_0 . It's easy to find $P(\text{Type I error})$
 - ii) Put the hypo you trust more on H_0 (with stronger evidence, not mentally)
 - iii) Put the hypo which will lead to serious consequence owing to reject it on H_0 .

e.g. put the hypo of some result of an experiment on H_1 . H_1 is also called research hypothesis.

Def: cutoff number z_α s.t. $P(Z \geq z_\alpha) = \alpha$.

(Distinguish it from quantile, percentile)

Def: A test with power function $\beta(\theta)$ is unbiased if $\beta(\theta') = \beta(\theta'')$, $\forall \theta' \in \Theta_0^c, \theta'' \in \Theta_0$.

② The most powerful test:

As before. we will restrict on a class C of tests for $H_0: \theta \in \Theta_0$, v.s. $H_1: \theta \in \Theta_0^c$, with level α . Then a good test in such class should have a small prob of Type II error.

\Rightarrow Def: A test in C with power function $\beta(\theta)$ is uniformly most powerful (UMP) if for any other test with $\beta'(\theta)$,

$$\Rightarrow \beta(\theta) \geq \beta'(\theta), \quad \forall \theta \in \Theta^c.$$

Remark: Specify the elements in C : (H_0 vs. H_1)

generalized randomized tests:

$$\phi(\vec{x}) = \begin{cases} 1, & \vec{x} \in R \\ 0, & \vec{x} \in R^c \\ \lambda, & \vec{x} \in \partial R \end{cases} \quad \begin{array}{l} (\lambda \text{ can be controlled} \\ \text{to set } \alpha \text{ size} \\ \text{test, exactly}) \end{array}$$

$$C = \{\phi_i \mid P(\phi_i = 1 \mid \theta \in \Theta_0) \leq \alpha\}.$$

To characterize UMP test, begin with both simple hypothesis (e.g. $\theta = \lambda$, for some const. λ)

Thm. (Neyman-Pearson Lemma)

Consider $H_0: \theta = \theta_0$ v.s. $H_1: \theta = \theta_1$, $\vec{X} \sim f(\vec{x}|\theta)$

For a test with rejection region R . St.

$\vec{x} \in R \Leftrightarrow f(\vec{x}|\theta_1) \geq k f(\vec{x}|\theta_0)$, for some $k \geq 0$.

(i.e. $\vec{x} \in R^c$ if $f(\vec{x}|\theta_1) < k f(\vec{x}|\theta_0)$).

Let $P_{\theta_0}(X \in R) = \alpha$.

i) Then it's the UMP test of α level.

ii) Conversely, every UMP test of α level (if exist)

satisfies the condition above, except a set A , where

$P_{\theta_0}(\vec{X} \in A) = P_{\theta_1}(\vec{X} \in A) = 0$. When there exists a test

satisfies the condition above, with replace " $k \geq 0$ " with " $k > 0$ ".

Pf: i) (Sufficient)

Denote that test ϕ . For any other test ϕ' with rejection region R' . St. $P_{\theta_0}(\vec{X} \in R') \leq \alpha$ → used to construct a UMP test

$$\text{Then } \beta(\theta_1) - \beta'(\theta_1) = \int_{R'} f(\vec{x}|\theta_1) - \int_{R'} f(\vec{x}|\theta_1)$$

$$= \int_{R \setminus R'} f(\vec{x}|\theta_1) - \int_{R \setminus R'} f(\vec{x}|\theta_1)$$

$$\geq k \left[\int_{R \setminus R'} f(\vec{x}|\theta_1) - \int_{R \setminus R'} f(\vec{x}|\theta_0) \right]$$

$$= k \left[\int_R f(\vec{x}|\theta_1) - \int_R f(\vec{x}|\theta_0) \right] \geq 0.$$

Replace " \int " with " Σ " for discrete case.

\Leftarrow (Necessary) Only for contr. case.

Note 3). " $=$ " holds when $P_{\theta_0}^L(R/k') = P_{\theta_0}^U(R/k) = 0$.

$\therefore \phi, \phi'$'s rejection region R, R' only differs
a prob. measure 0. set.

→ Used to
show nonexistence
of UMP test.
(Because $\exists R$
has some kind
of Uniqueness)

Remark: i) The point is that it transforms "Comparing power function (Φ)" to "Comparing level" by the condition of rejection region.

ii) For $T(\bar{X})$ s.s. for θ , the condition can be relaxed: (suppose test based on T has R_T = reject. region)

$$t \in R_T \Leftrightarrow g(t|\theta_1) \geq k g(t|\theta_0)$$

Next, we consider one of hypothesis is composite hypo:

Def: A family of pdf's or pmf's $\{g(t|\theta) | \theta \in \Theta\}$. for univariate r.v. T has a monotone likelihood Ratio (MLR) if for $\forall \theta_2 > \theta_1$, $\frac{g(t|\theta_2)}{g(t|\theta_1)}$ is mon. on t .

Remark: $X \sim h(t)(\theta) e^{w(\theta)t}$. w is monotone. Then, it has MLR!

Thm. (Karlin-Rubin)

Consider $H_0: \theta \leq \theta_0$ v.s. $H_1: \theta > \theta_0$. $T \sim g(t|\theta)$ which has a MLR of increasing. T is s.s. for θ . Then $\forall t_0$, " $\text{Rejects } H_0 \text{ when } T \geq t_0$ " is UMP $\alpha = P_{\theta_0}(R|H_0)$ test.

Pf: For returning to simple hypothesis

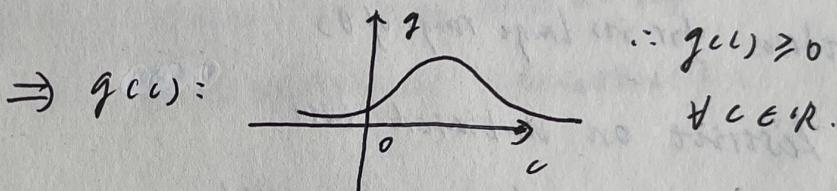
Fix $\theta' > \theta_0$. Consider $H_0: \theta = \theta_0$, v.s. $H_1: \theta = \theta'$.

Lemma. If $g(t|\theta)$ of T has an increasing type MLR. Then for $\theta_1 < \theta_2$, $P_{\theta_1} (T > c) \leq P_{\theta_2} (T > c)$, H.C.

Pf: Let $g(c) = P_{\theta_2} (T > c) - P_{\theta_1} (T > c)$

$$g'(c) = g(c|\theta_1) - g(c|\theta_2) = g(c|\theta_1) \left(1 - \frac{g(c|\theta_2)}{g(c|\theta_1)}\right)$$

Note $g(c|\theta_2)/g(c|\theta_1) \uparrow$. $g(c|\infty) = g(-\infty) = 0$.



$\Rightarrow \beta(\theta)$ is mondecreasing on θ .

i) $\sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha$. So it's level of test.

ii) Def: $k' = \inf_{t \in T} \frac{g(t|\theta')}{g(t|\theta_0)}$, $T = \{t > t_0, g(t|\theta') \text{ or } g(t|\theta_0) > 0\}$.

$\therefore T \geq t_0 \Leftrightarrow g(t|\theta') \geq k' g(t|\theta_0) \Leftrightarrow t \in T$.

\therefore For any other α level test with $\beta^*(\theta)$.

Since $\beta^*(\theta_0) = \sup_{\theta \leq \theta_0} \beta(\theta) \leq \alpha$. $\therefore \beta^*(\theta') \leq \beta(\theta')$. $\forall \theta' > \theta_0$.

Since θ' is arbitrary. It holds for $H_0: \theta = \theta_0$ v.s. $H_1: \theta > \theta_0$.

By ii). extend to $H_0: \theta \leq \theta_0$ v.s. $H_1: \theta > \theta_0$.

Remark: i) For decreasing type. Let $T = -T$.

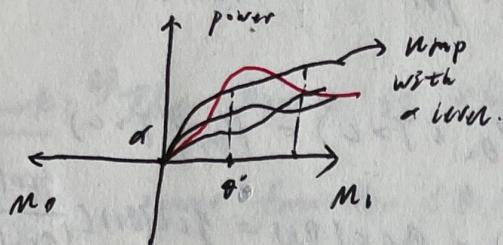
reverse $\lambda: T < t_0$.

ii) For $H_0: \theta \geq \theta_0$, v.s. $H_1: \theta < \theta_0$. Similarly.

reverse $T \leq t_0$!

iii) We can operate hypothesis from: simple v.s. simple.

\Rightarrow composite v.s. simple \Rightarrow composite v.s. composite.



(If the red curve exists.

Then UMP test with α level
doesn't exist.)

(Nonexistence of UMP test
often shows up in two sided
hypotheses for very large range of θ)

\Rightarrow restrict on unbiased test

may result in finding UMP test!

③ Size of NI and

IU tests :

NI Test :

A relationship between LRT and NI:

Thm. For $H_0: \theta \in \Theta_0$ v.s. $H_1: \theta \in \Theta_0^c$, $\Theta_0 = \bigcap_{y \in I} \Theta_y$

$\lambda_y(\vec{x})$ is LRT for each H_0 v.s. H_1 , $\lambda(x)$

is LRT for H_0 v.s. H_1 , $T(x) = \inf_{y \in I} \lambda_y(x)$

For test one: $R: [T(x) \leq c]$ with $\beta_T(\theta)$

test two: $R: [\lambda(x) \leq c]$ with $\beta_\lambda(\theta)$.

Then, $T(\vec{x}) \geq \lambda(\vec{x}) \quad \forall \vec{x} \in \Omega$. Besides,

$\beta_T(\theta) = \beta_\lambda(\theta)$, if $\theta \in \Theta$. So if LRT test is level α , then NLT is α level.

Pf: $\forall \gamma \in I$. $\lambda_\gamma(\vec{x}) \geq \lambda(\vec{x})$, since $\Theta_\gamma \supseteq \Theta$.

From $T(\vec{x}) \geq \lambda(\vec{x}) \Rightarrow \{T(\vec{x}) \leq c\} \subseteq \{\lambda(\vec{x}) \leq c\}$.

$\therefore \beta_T(\theta) \leq \beta_\lambda(\theta)$, $\because \sup_{\theta \in \Theta} \beta_\lambda(\theta) \leq \sup_{\theta \in \Theta} \beta_T(\theta) < \gamma$.

Remark: LRT is more powerful than NLT. But we usually use NLT:

- { i) NLT has smaller type I error prob.
- { ii) If H_0 is rejected, we can look at H_0 to see why! (additional information!)

ii) IN Test:

Thm. α_Y is size of test of H_0 v.s. H_A with rejection region R_Y . Then INT with $R = \bigcap_{Y \in I} R_Y$ is level

$\alpha = \sup_{Y \in I} \alpha_Y$ test.

Pf: $\forall \theta \in \Theta$. $P_\theta(X \in R) \leq P_\theta(X \in R_Y)$, since $R \subseteq R_Y$, $\forall Y \in I$.

Remark: The shortcoming is that it's conservative. Since α may be much larger than its size.

Thm. For $H_0: \theta \in \bigcup_i^k \Theta_i$, R_i is rejection region for $H_{0i}: \theta \in \Theta_i$ of level α . If for some i , $1 \leq i \leq k$, st. exists a seq of parameters $\{\theta_n\} \subseteq \Theta_i$, st.

$$\begin{cases} \lim_{n \rightarrow \infty} P_{\theta_n}(X \in R_i) = 1 \\ \forall j \neq i, \lim_{n \rightarrow \infty} P_{\theta_n}(X \in R_j) = 1 \end{cases} \Rightarrow \text{INT with } R = \bigcap_i^k R_i \text{ is size-}\alpha \text{ test.}$$

Pf. By Thm above. $\sup_{\theta \in \Theta_0} P_\theta(\vec{X} \in R) \leq \alpha$.

$$\sup_{\theta \in \Theta_0} P_\theta(\vec{X} \in R) \geq \lim_{l \rightarrow \infty} P_{\theta_l}(\vec{X} \in R) = \lim_{l \rightarrow \infty} P_{\theta_l}(\vec{X} \in \bigcap_{i=1}^k R_i)$$

$$\geq \lim_{l \rightarrow \infty} \sum_{i=1}^k P_{\theta_l}(X \in R_i) - (k\alpha) = \alpha.$$

$$\text{Since } P_{\theta_l}(\vec{X} \in \bigcap_{i=1}^k R_i) = 1 - P_{\theta_l}(\vec{X} \in \bigcup_{i=1}^k R_i^c)$$

$$\geq 1 - \sum_{i=1}^k P_{\theta_l}(\vec{X} \in R_i^c) = \sum_{i=1}^k P_{\theta_l}(\vec{X} \in R_i^c) - (k\alpha) \quad \square$$

Remark: Particular case: $\sup_{\theta \in \Theta_0} \beta_i(\theta) = \lim_{l \rightarrow \infty} P_{\theta_l}(\vec{X} \in R_i) = \alpha$.

④ P-value:

→ α hasn't given out yet.

Then we can determine another way of reporting the result (Not use Accept or Reject) by giving "α" of test is to report a statistic — p-value.

Def: p-value $p(\vec{x})$ is a test statistic. st. $0 \leq p(\vec{x}) \leq 1$.

& $\vec{x} \in R$. small $p(\vec{x})$ means it's skewed. giving evidence that H_0 is true.

(*) Because: ←

$P_\theta(p(X) \leq \alpha) \Rightarrow$ A p-value is valid if $P_\theta(p(\vec{x}) \leq \alpha) \leq \alpha$. (*)

$= P_\theta(X \in R)$. So use it to construct R . we can obtain a level α test.

The most common way to define a p-value:

Thm. $W(\vec{x})$ is a test statistic. st. large value of $W(\vec{x})$ gives evidence to H_0 is true. Def.

$$p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{x}) \geq W(\vec{x}_0)), \text{ for each } \vec{x}.$$

Then $p(\vec{x})$ is valid p-value.

$$\underline{\text{Pf:}} \quad P_\theta(W(\vec{x}) \geq w(\vec{x})) = P_\theta(-W(\vec{x}) \leq -w(\vec{x}))$$

$$= F_\theta(-W(\vec{x})). = P_\theta(\vec{x}). \quad F_\theta \text{ is cdf of } -W(\vec{x}).$$

$$\therefore P_\theta(p(\vec{x}) \leq \tau) \leq P_\theta(P_\theta(\vec{x}) \leq \tau) = P_\theta(F_\theta(-W(\vec{x})) \leq \tau)$$

$$= P_\theta(-W(\vec{x}) \in A_\tau) = F_\theta(-W(\vec{x})) \leq \tau.$$

$$A_\tau = \{ -W(\vec{x}) \mid F_\theta(-W(\vec{x})) \leq \tau \} = (-\infty, -W(\vec{x})]. \quad (\text{right-closed})$$

where $F_\theta(-W(\vec{x})) \leq \tau$. Since $-W(\vec{x}) \in A_\tau$.

Remark: i) $p(\vec{x}) \downarrow$ means $W(\vec{x}) > W(\vec{x})$ has small prob.

$\Leftrightarrow W(\vec{x})$ is large. \therefore gives support to H_1 .

ii) An useful Lemma:

• X has cdf $F_X(\cdot)$. Then $p(F_X(x) \leq x) = x$.

Pf: $\{X \mid F_X(x) \leq x\} = (-\infty, tx] \text{ or } (-\infty, tx) = A_x$.

The second case can happen in discrete case.

$\therefore \lim_{t \uparrow tx} F_X(t) = F_X(tx) \leq x. \quad (\text{By continuity})$

$\therefore p(F_X(x) \leq x) = p(X \in A_x) = p(X \leq tx) =$

$F_X(tx) \leq x. \quad \text{Since } tx \in A_x.$

iii) Another interpretation of p-value: (Observe significant level)

Under the condition in the Thm.

If c_α is a critical value chosen s.t. $\{x \mid W(x) \geq c_\alpha\}$

is a rejection region of size α test of H_0 . Then

$p(\vec{x})$ is the smallest value of level that rejects H_0 .

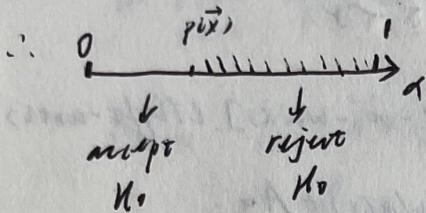
Pf: $\tau = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{x}) \geq c_\alpha)$

$p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{x}) \geq W(\vec{x}))$

$$\text{If } p(\vec{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{X}) \geq W(\vec{x})) \geq \alpha = \sup_{\theta \in \Theta_0} P_\theta(W(\vec{X}) \geq c_\alpha)$$

$\therefore W(\vec{x}) \geq c_\alpha$. (Argue by contradiction). Reject H_0

Conversely if $p(\vec{x}) < \alpha$, we have $W(\vec{x}) < c_\alpha$. accept H_0



when $p(\vec{x})$ is extreme, we need to support H_0 strongly to accept H_0 .

Remark: ii) Note that the smaller p-value is, the more chance we can reject H_0 . That's because if we want to accept H_0 , then the prob. of type I error should be small enough ($< p\text{-value}$)
 iii) p-value is the "level" of experiment we have done.

Another method to

construct p-value:

Suppose $S(\vec{X})$ is a.s.s. for $\{f(\vec{x}|\theta) / \theta \in \Theta_0\}$

and $\{f(\vec{x}|\theta) / \theta \in \Theta\}$, $\Theta_0 \subseteq \Theta$. Then for each $\vec{x} \in \mathcal{X}$,

Dof: $p(\vec{x}) = P_{\theta^*}(W(\vec{X}) \geq W(\vec{x}) | S = S(\vec{x}))$ ($\text{indpt of } \theta \in \Theta_0$)

Then it's valid: for $\forall \theta \in \Theta_0$

$$P_{\theta^*}(p(\vec{X}) \leq \alpha) = \sum_S P_{\theta^*}(p(\vec{X}) \leq \alpha | S = S) P_{\theta^*}(S = S)$$

$$\leq \sum_S \alpha P_{\theta^*}(S = S) = \alpha.$$

Remark: p-value for IUT:

$H_0: \theta \in \bigcup_{j=1}^k \Theta_j$. P_j is p-value for $H_0|j$. Then

$p(\vec{x}) = \max_{1 \leq j \leq k} P_j(\vec{x})$ is a valid p-value for H_0 .

$$(P_{\theta^*}(p(\vec{X}) \leq \alpha) \leq P_{\theta^*}(P_j(\vec{x}) \leq \alpha) \leq \alpha, \text{ since } \exists j)$$

$\theta \in \Theta_j$. $P_j(\vec{x}) \leq p(\vec{x})$. Analogous for NUT!