STA355 # Lecture 7

Hypothesis testing. $\Theta_0 \cap \Theta_1 = \phi$.

Decide: $H_0: \theta \in \Theta_0$ v.s. $H_1: \theta \in \Theta_1$

based on observed data.

Usually, we compute some test statistics $T(X) \in \mathbb{R}$

and reject when $T(X) > c$

for some pre-specified (deterministic) $c \in \mathbb{R}$.

"critical value".

Type-I error: $\theta \in \Theta_0$, but reject

Type-II error: $\theta \in \Theta_1$, but cannot reject.

Power function $\beta(\theta) = \mathbb{P}_\theta(\text{reject})$ $(\forall \theta \in \Theta)$.

size/level of the test

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

"worst-case type-I error"

Goal of testing

$$\text{maximize } \{ \beta(\theta) \text{ for } \theta \in \Theta_1 \}$$

Still a multi-objective optimization problem.

$$\text{s.t.} \quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha_0 \quad (\text{e.g. } 5\%).$$

A valid test: type-I error constraint is satisfied.

(Asymptotic version: $X_1, X_2, \cdots, X_n \overset{iid}{\sim} \mathbb{P}_\theta$.

$$\text{level} = \sup_{\theta \in \Theta_0} \lim_{n \to +\infty} \mathbb{P}_\theta(\text{reject})$$

Concrete instantiation:

1-dim problem:

- Two-sided testing:

$$H_0: \theta = \theta_0 \quad \text{v.s.} \quad H_1: \theta \ne \theta_0$$

(need to ensure $\mathbb{P}_{\theta_0}(\text{reject}) \le \alpha$).

- One-sided testing

$$H_0: \theta \le \theta_0 \quad \text{v.s.} \quad H_1: \theta > \theta_0$$

(or the other way around)

Need to ensure $\forall \theta \le \theta_0$, $\mathbb{P}_\theta(\text{reject}) \le \alpha$

e.g. $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$

$$T(X) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

- One-sided testing: $H_0: \theta \le \theta_0$, $H_1: \theta > \theta_0$.

Reject when $T(X) > c$

under θ, $T(X) \sim N(\theta, \frac{1}{n})$

$$\mathbb{P}_\theta(\text{reject}) = 1 - \Phi(\sqrt{n}(c-\theta))$$

(where Φ is cdf of $N(0,1)$).

So we choose $c = \theta_0 + \frac{1}{\sqrt{n}} Z_{1-\alpha}$

($(1-\alpha)$ quantile of $N(0,1)$).

Similarly, for two-sided tests.

reject when $T(X) > c_2$ or $T(X) < c_1$

$$\mathbb{P}_{\theta_0}(\text{rejection}) = 1 - \Phi(\sqrt{n}(c_2 - \theta_0)) + \Phi(\sqrt{n}(\theta_0 - c_1)).$$

Need to choose (c_1, c_2) pair, s.t. $\mathbb{P}_{\theta_0}(\text{rejection}) \le \alpha$.

- Commonly used choice:

$$\begin{cases} c_2 = \theta_0 + \frac{1}{\sqrt{n}} z_{1-\alpha/2} \\ c_1 = \theta_0 - \frac{1}{\sqrt{n}} z_{1-\alpha/2} \end{cases}$$

"unbiased test"

~~most~~ powerful among unbiased tests.

- For any $\gamma \in (0,1)$.

$$\begin{cases} c_2 = \theta_0 + \frac{1}{\sqrt{n}} z_{1-\gamma\alpha} \\ c_1 = \theta_0 - \frac{1}{\sqrt{n}} z_{1-(1-\gamma)\alpha} \end{cases}$$

is also a valid choice.

Wald test:

- $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

Suppose that we have estimator $\hat{\theta}_n$

s.t. under $\mathbb{P}_{\theta_0}$, we have $\frac{\hat{\theta}_n - \theta_0}{\widehat{se}} \xrightarrow{d} N(0,1)$.

then we reject when

$$\left|\frac{\hat{\theta}_n - \theta_0}{\widehat{se}}\right| \geq z_{1-\alpha/2}$$

By def of "$\xrightarrow{d}$", we have

$$\lim_{n\to+\infty} \mathbb{P}_{\theta_0}(\text{rejection}) = \alpha.$$

Similarly, we have one-sided test:

$H_0: \theta \leq \theta_0$, $H_1: \theta > \theta_0$

reject when $\frac{\hat{\theta}_n - \theta_0}{\widehat{se}} > z_{1-\alpha}$.

χ^2 tests.

Suppose that $\theta \in \mathbb{R}^d$

$H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

Estimator $\hat{\theta}_n$: (under $\mathbb{P}_{\theta_0}$) $(\hat{\Sigma})^{-1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_d)$

($\hat{\Sigma}$ computed from data, estimator for asymp cov).

If $\xi \sim \mathcal{N}(0, I_d)$, then $\|\xi\|_2^2 \sim \chi_d^2$

χ^2 test: reject when

$$\|\hat{\Sigma}_n^{-1/2}(\hat{\theta}_n - \theta_0)\|_2^2 \geq (1-\alpha) \text{ quantile of } \chi_d^2.$$

Remark: $\hat{se}$ and $\hat{\Sigma}$ can also be replaced by exact expressions of asymp. var. under $\mathbb{P}_{\theta_0}$.

The results will still be valid.

(we use $\hat{se}$ and $\hat{\Sigma}$ for computational reasons).

Special case:

$X \sim \text{Multinomial}(n, p)$ for $p \in \mathbb{R}^k$

$p_1, p_2 \cdots p_k > 0$

$\sum_{j=1}^k p_j = 1$

Want to test

$H_0: p = p_0$

$H_1: p \neq p_0$.

Under p_0.

$$\sqrt{n}\left(\frac{X}{n} - p_0\right) \xrightarrow{d} \text{degenerate normal.}$$

Pearson's χ^2 test.

under H_0

$$T = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} \xrightarrow{d} \chi^2_{k-1}$$

Student t-test.

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

μ, σ^2 are unknown, we want to test

$H_0: \mu = \mu_0$, vs. $H_1: \mu \neq \mu_0$.

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\widehat{se}_n} \quad \left(\xrightarrow{d} N(0,1), \text{ as } n \to +\infty\right)$$

"t-distribution w/ dof $n-1$"

$$:= \text{distribution of } \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i}{\widehat{se}(\xi_1, \cdots, \xi_n)}$$

where $\xi_1, \xi_2, \cdots, \xi_n \overset{iid}{\sim} N(0,1)$.

reject based on quantile of t-distribution

$\Rightarrow$ exact level-α test.

p-values.

Def n. Suppose $T(X)$ is the test statistics.

Rejection region R_α (reject when $T(X) \in R_\alpha$).

p-value $:= \inf\{\alpha : T(X) \in R_\alpha\}$.

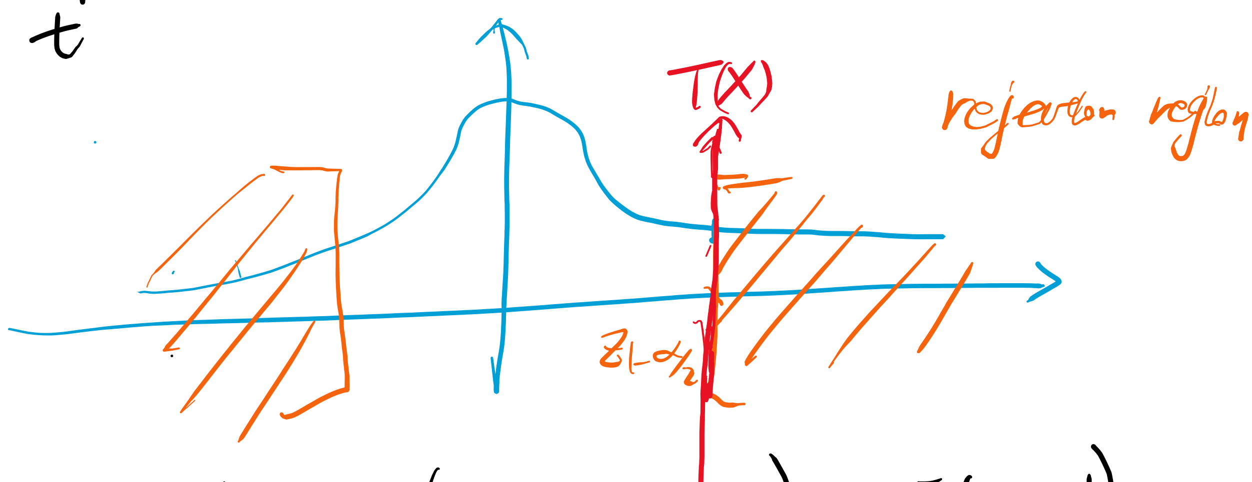
(read the warnings from textbook)

Test becomes: reject when p-value $\leq \alpha$.

e.g. $T(X) = \dfrac{\hat{\theta} - \theta_0}{\hat{se}}$. $R_\alpha := \{t \in \mathbb{R} : |t| > z_{1-\alpha/2}\}$.

$\parallel$

t



$$\text{p-value}(t) = \mathbb{P}_{\theta_0}(|T(X)| \geq |t|) = 2\Phi(-|t|)$$

So the p-value is given by $2\Phi(-|T(X)|)$.

(in other words, we can solve for p-value by

$$z_{1-\alpha/2} = T(X)\Big]$$

Fact: If $T(X)$ has a cts distribution under $\mathbb{P}_{\theta_0}$

then the p-value $\sim \mathrm{Unif}([0,1])$ under $\mathbb{P}_{\theta_0}$.

(This is a direct corollary of the fact:

If X has cts distribution, F is its cdf

then $F(X) \sim \mathrm{Unif}([0,1])$)

Multiple testing.

A collection of testing problems

H_{0i} vs. H_{1i} for $i=1,2,\cdots,m$.

$P_1, P_2, \cdots, P_m$ are p-values computed

for each testing problem

Naive idea: reject P_j when $P_j \le \alpha$

for each $j \in \{1,2,\cdots,m\}$.

e.g. If P_j's are all independent, and test stats have cts distributions.

$$\mathbb{P}_{H_0}(\text{make any false discovery}) = 1-(1-\alpha)^m$$

when $m \approx \frac{1}{\alpha}$, $\mathbb{P}(\cdots) \approx 1-\frac{1}{e}$

Bonferroni correction.

Idea: (no matter what dependence structures)

$$\mathbb{P}_{H_0}(\text{false discovery}) \leq \sum_{j=1}^{m} \mathbb{P}_{H_{0j}}(\text{reject } H_{0j}) \leq \alpha$$

We choose to test each H_{0j} at level $\frac{\alpha}{m}$.

Relaxation: False Discovery Rate (FDR).

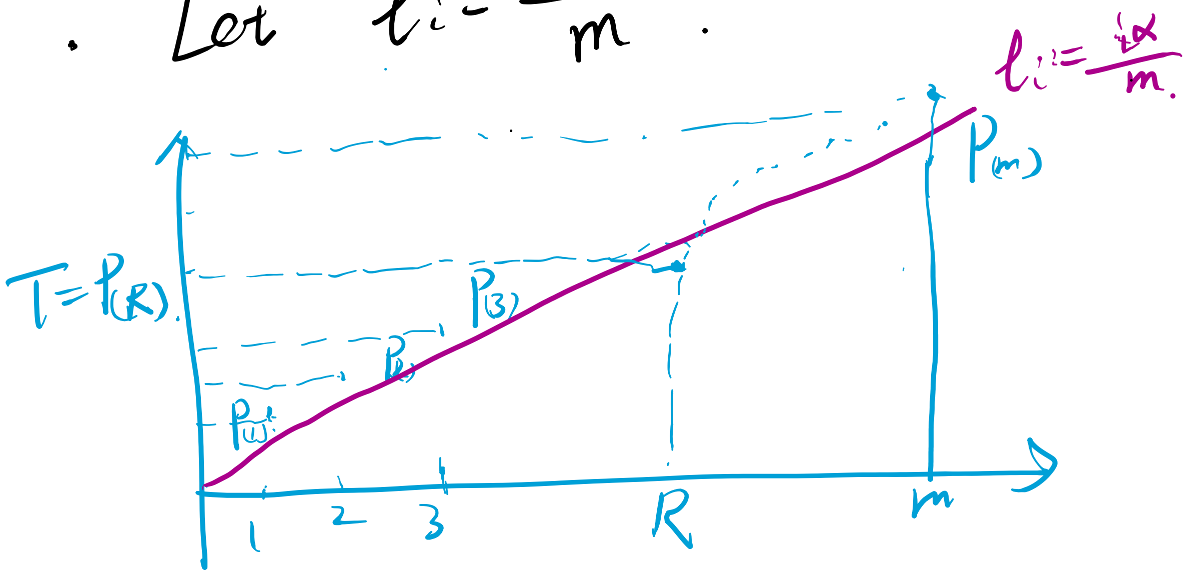
False discovery proportion.

$$\mathrm{FDP} := \begin{cases} \dfrac{\#\text{ false rejections}}{\#\text{ rejections}} & (\#\text{rej} > 0) \\ 0 & (\#\text{rej} = 0) \end{cases}$$

$$\mathrm{FDR} = \mathbb{E}[\mathrm{FDP}].$$

Benjamini-Hochberg (BH) procedure:

- $P_{(1)} < P_{(2)} \cdots < P_{(m)}$ (sorted p-values).
- Let $l_i := \frac{i\alpha}{m}$.



- $R := \max\{i : P_{(i)} \leq l_i\}$.

We let $T = P_{(R)}$.

and we reject when $P_i \leq T$

(i.e. we reject the R smallest p-values)

Permutation tests.

Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} P_X$

$Y_1, Y_2, \dots, Y_m \overset{iid}{\sim} P_Y$

$H_0: P_X = P_Y \qquad H_1: P_X \neq P_Y$.

Idea: under H_0, $(n+m)$ samples can be mixed together, permutation does not make much difference

Given a test stats

$T(X_1, X_2, \dots X_n, Y_1, Y_2, \dots Y_m)$

(e.g. $\left|\frac{1}{n}\sum_{1}^{n} X_i - \frac{1}{m}\sum_{1}^{m} Y_j\right|$,

$\sup_{g \in G} \left|\frac{1}{n}\sum_{1}^{n} g(X_i) - \frac{1}{m}\sum_{1}^{m} g(Y_i)\right|$).

Let $N = n+m$, consider all possible $N!$ permutations of $(X_1, \dots, X_n, Y_1, \dots Y_m)$.

(for notation convenience, we let $X_{n+i} = Y_i$
$i=1, \dots m$)

For each permutation σ

Let

$$T_\sigma := \text{test statistic computed on}$$

$$\underbrace{X_{\sigma(1)}, X_{\sigma(2)}, \cdots X_{\sigma(n)}}_{\text{as } X's}, \underbrace{X_{\sigma(n+1)} \cdots X_{\sigma(n+m)}}_{\text{as } Y's.}$$

$$p\text{-value} := \frac{1}{N!} \sum_{\sigma} \mathbb{1}\{T_\sigma > T(X;Y)\}.$$

(Percentile of observed $T(X;Y)$ among $N!$ hypothetical worlds)

Monte-Carlo simulation: (similar to bootstrap)

For $b = 1, 2, \cdots B$,

- sample σ_b uniformly among permutations.

- Compute T_{σ_b}.

$$p\text{-value} \approx \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{T_{\sigma_b} > T(X;Y)\}.$$