Outline

1) Hypothesis testing
2) Neyman - Pearson Lemma
3) Uniformly most powerful tests

Hypothesis Testing

Model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$

Null hypothesis $H_0 : \theta \in \Theta_0$ ("default")

Alternative hyp. $H_1 : \theta \in \Theta_1$

Hypotheses should be disjoint : $\Theta_0 \cap \Theta_1 = \emptyset$

and exhaustive $\Theta_0 \cup \Theta_1 = \Theta$

Want to use data X to learn which includes θ

Inductive behavior:

We either reject H_0 (conclude $\theta \in \Theta_1$), or fail to reject H_0 (no conclusion)

("Accept H_0" ok as technical term, but may confuse)

$H_{0/1}$ called simple if $\Theta_{0/1} = \{\theta_{0/1}\}$ composite o.w.

Ex $X \sim N(\theta, 1)$

$H_0 : \theta \le 0$ vs $H_1 : \theta > 0$ (composite vs. composite)

$H_0 : \theta = 0$ vs $H_1 : \theta \ne 0$ (simple vs composite)

Ex $X_1, \ldots, X_n \sim P$ $Y_1, \ldots, Y_m \sim Q$

$H_0 : P = Q$ vs $H_1 : P \ne Q$ (composite vs. composite)

Critical Function

Can describe a test formally by its **critical function** (a.k.a. **test function**)

$$\phi(x) = \begin{cases} 0 & \text{accept } H_0 \\ \pi \in (0,1) & \text{reject w.p. } \pi \\ 1 & \text{reject } H_0 \end{cases}$$

In practice, randomization rarely used $\left(\phi(x) = \{0,1\}\right)$
(In theory, simplifies discussions.)

A non-randomized test partitions $\mathcal{X}$ into

$R = \{x : \phi(x) = 1\}$ **rejection region**

$A = \{x : \phi(x) = 0\}$ **acceptance region**

Usually defined via **test statistic** $T(X) \in \mathbb{R}$

We say ϕ **rejects for large $T(X)$** if

$$\phi(X) = \begin{cases} 0 & T(X) < c \\ 1 & T(X) > c \\ \gamma \in (0,1) & T(X) = c \quad \text{(if randomized)} \end{cases}$$

for **critical threshold** $c \in \mathbb{R}$

T chosen to discriminate well between H_0, H_1

Significance Level and Power

Two types of errors

1) Type I error : H_0 true but we reject
2) Type II error : H_0 false but we don't rej.

Usual goal is to minimize $\mathbb{P}_{H_1}(\text{Type II error})$,
while controlling $\mathbb{P}_{H_0}(\text{Type I error}) \leq$ fixed α

Note if $H_{0/1}$ composite, "$\mathbb{P}_{H_{0/1}}$" is not a well-defined prob.

<u>Power function</u> :

$$\beta_\phi(\theta) = \mathbb{E}_\theta[\phi(X)]$$
$$= \mathbb{P}_\theta[\text{Reject } H_0]$$

fully summarizes test's behavior

<u>Goal</u> :

multiple objectives ... multiple constraints

maximize β_ϕ for $\theta \in \Theta_1$, subject to $\beta_\phi \leq \alpha$ for $\theta \in \Theta_0$

ϕ is a <u>level-α test</u> ($\alpha \in [0,1]$) if $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \alpha$

Ubiquitous choice is $\alpha = 0.05$

["Most influential offhand remark in history of science"]

<u>Question</u>: Can we find ϕ^* that maximizes power <u>everywhere</u> on the alternative at once?

Z test

Ex Test statistic $Z(X) \sim N(\theta, 1)$ (very common)

Def Upper α quantile $z_\alpha = \Phi^{-1}(1-\alpha)$, $\Phi = N(0,1)$ cdf

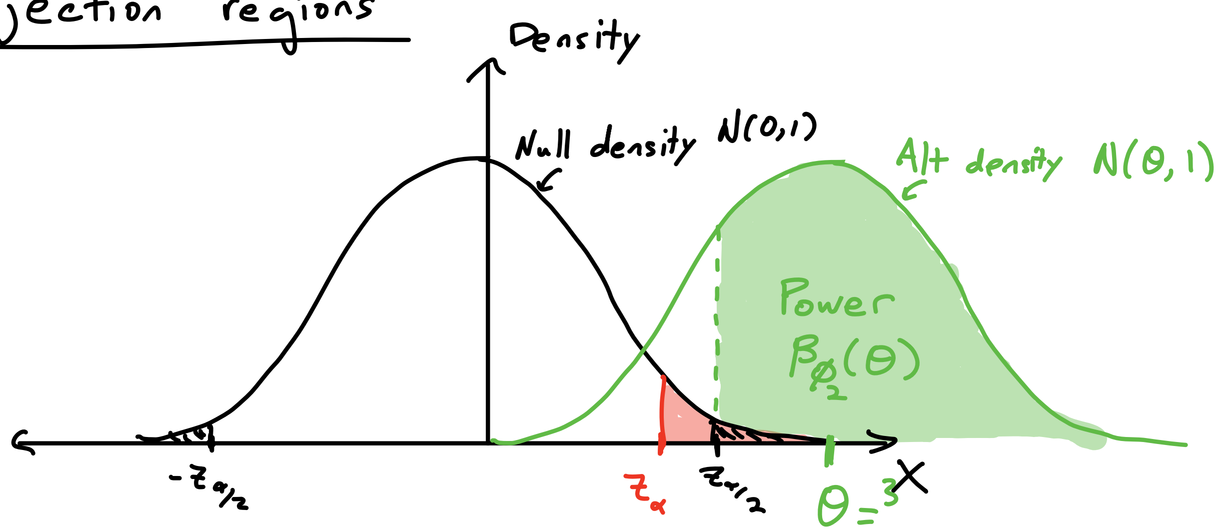
1-sided Z-test : $H_0: \theta \le 0$ vs $H_1: \theta > 0$

$\phi_1(x) = 1\{x > z_\alpha\}$

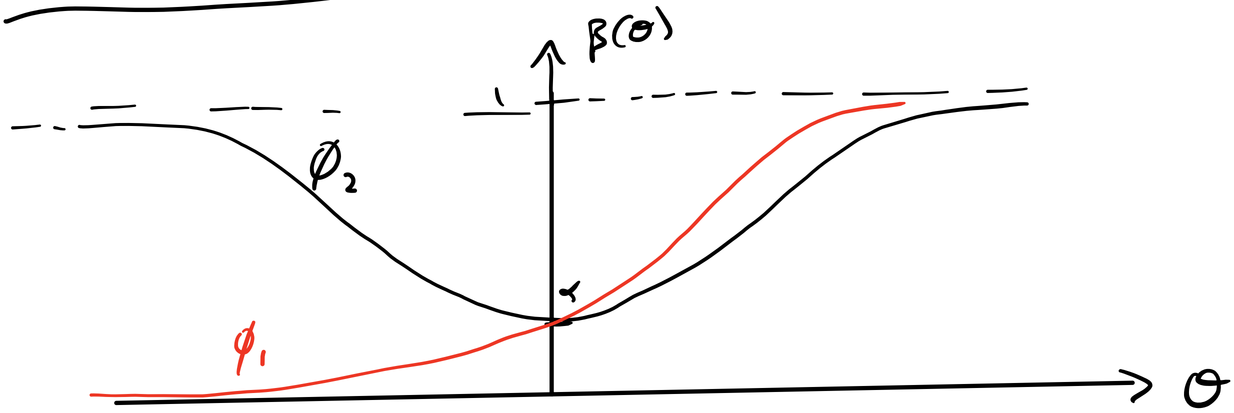
2-sided Z-test : $H_0: \theta = 0$ vs $H_1: \theta \neq 0$

$\phi_2(x) = 1\{|x| > z_{\alpha/2}\}$ (Could also use $\phi_1(x)$)

Rejection regions



Power functions:



Likelihood Ratio Test

Simple vs simple: $H_0: X \sim P_0$ vs $H_1: X \sim P_1$

Densities p_0, p_1 wrt dominating measure μ (e.g. $P_0 + P_1$)

Optimal test rejects for large values of

Likelihood ratio: $LR(x) = p_1(x)/p_0(x)$

Likelihood ratio test (LRT):

$$\phi^*(x) = \begin{cases} 1 & LR(x) > c \\ \gamma & LR(x) = c \\ 0 & LR(x) < c \end{cases}$$

c, γ chosen to make $\mathbb{E}_0 \phi^*(x) = \alpha$

Intuition:

Power under H_1: $\max \int_R p_1(x)\,d\mu(x)$ "Bang"

Sig. budget: $\int_R p_0(x)\,d\mu(x) \le \alpha$ "Buck"

Spend fixed α budget on x values that deliver greatest bang/buck

Neyman-Pearson

Theorem (Neyman – Pearson Lemma)

LRT with significance level α is optimal for testing $H_0: X \sim p_0$ vs. $H_1: X \sim p_1$

Proof We are interested in maximization problem

$$\underset{\phi: \mathcal{X} \to [0,1]}{\text{maximize}} \quad \mathbb{E}_1[\phi(X)] \quad \text{s.t.} \quad \mathbb{E}_0[\phi(X)] \leq \alpha$$

Lagrange form:

$$\text{maximize} \quad \mathbb{E}_1[\phi(x)] - \lambda \mathbb{E}_0[\phi(X)]$$

$$= \int \phi(x) \left(p_1(x) - \lambda p_0(x) \right) d\mu(x)$$

$$= \int \phi(x) \left(\frac{p_1(x)}{p_0(x)} - \lambda \right) dP_0(x)$$

Solution(s):

$$\phi(x) = \begin{cases} 1 & \text{if } LR > \lambda \\ 0 & \text{if } LR < \lambda \\ \text{arbitrary} & \text{if } LR = \lambda \end{cases}$$

$\Rightarrow \phi^*$ maximizes Lagrangian for $\lambda = c$

consider any other test $\tilde{\phi}(x)$, $\mathbb{E}_0 \tilde{\phi}(x) \leq \alpha$

$$\mathbb{E}_1 \tilde{\phi} \leq \mathbb{E}_1 \tilde{\phi} - c\mathbb{E}_0 \tilde{\phi} + c\alpha \qquad c(\alpha - \mathbb{E}_0 \tilde{\phi}) \geq 0$$

$$\leq \mathbb{E}_1 \phi^* + c\,\mathbb{E}_0 \phi^* + c\alpha \qquad \phi^* \text{ maxes Lagrangian}$$

$$\leq \mathbb{E}_1 \phi^* \qquad c(\alpha - \mathbb{E}_0 \phi^*) \geq 0$$

Ex $X \sim \text{Binom}(n, \theta)$

Test $H_0 : \theta = .5$ vs $H_1 : \theta = .51$ at level $\alpha = 0.05$

$$p_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$\Rightarrow \quad \frac{p_{.51}(x)}{p_{.5}(x)} = \frac{.51^x (.49)^{n-x}}{.5^n} \propto_x \left(.51/.49 \right)^x \nearrow \text{ in } x$$

Reject for large $LR(X) \Leftrightarrow$ Rej. for large X

$\rightsquigarrow$ test stat X,

threshold $c_\alpha = 95^{th}$ %ile of $\text{Binom}(n, .5)$

X discrete: $\mathbb{P}_{H_0}(X > c_\alpha) < \alpha$ (since $2^{-n} \nmid .05$)

Randomize to "top off" error budget:

$$\text{Set } \gamma = \frac{\alpha - \mathbb{P}(X > c_\alpha)}{\mathbb{P}_{H_0}(X = c_\alpha)} \rightsquigarrow \mathbb{P}_{H_0}(\text{Reject } H_0) = \alpha$$

In practice just reject for $X > c_\alpha$

<u>conservative</u> test (sig. level $< \alpha$)

What about testing $H_0 : \theta = 0.5$ vs $H_1 : \theta = 0.508$?

Same test! $\left(\frac{.508}{.492} \right)^x$ also $\nearrow$ in X

Uniformly most powerful tests

General H_0, H_1 (simple or composite)

Def If $\phi^*(x)$ is level-α and for any other level-α test ϕ we have

$$\mathbb{E}_\theta \phi^* \geq \mathbb{E}_\theta \phi \qquad \forall \theta \in \Theta,$$

then ϕ^* is uniformly most powerful (UMP)

Def Assume $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ has densities p_θ.

$\mathcal{P}$ has monotone likelihood ratios (MLR) in $T(x)$

if $p_{\theta_2}(x) / p_{\theta_1}(x)$ is non-decreasing function of $T(x)$,

whenever $\theta_1 < \theta_2$ ($\frac{c}{0} := \infty$ if $c > 0$, $\frac{0}{0}$ undefined)

Ex. 1-param exponential family

$$X_1, \ldots, X_n \overset{iid}{\sim} p_\eta(x) = e^{\eta T(x) - A(\eta)} h(x)$$

$$p_{\eta_1}(x) / p_{\eta_0}(x) = \exp\{(\eta_1 - \eta_0) \sum_i T(x_i) - n(A(\eta_1) - A(\eta_0))\}$$

$$\nearrow \text{ in } T(X) = \sum_i T(x_i)$$

$\Rightarrow$ LRT rejects for large T if $\eta_1 > \eta_0$

" small T if $\eta_1 < \eta_0$

__Theorem__ Assume $\mathcal{P}$ has MLR in $T(X)$, and consider testing $H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$, for $\theta_0 \in \Theta \subseteq \mathbb{R}$

If $\phi^*(x)$ rejects for large $T(x)$,

ϕ^* is UMP at level $\alpha = \mathbb{E}_{\theta_0} \phi^*(X)$

__Proof__

Consider any other level-α test ϕ, any $\theta_1 > \theta_0$

ϕ is level-α for $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$

$p_{\theta_1}(x) / p_{\theta_0}(x)$ non-decr in $T(x)$ by assumption

$\Rightarrow \phi^*$ is LRT, $\beta_{\phi^*}(\theta_1) \ge \beta_\phi(\theta_1)$

Note $\beta_{\phi^*}(\theta_1) \ge \alpha$ for $\theta_1 > \theta_0$ (compare to $\phi(x) \equiv \alpha$)

Remains to show $\beta_{\phi^*}(\theta) \le \alpha$ for $\theta < \theta_0$

Consider testing $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ for $\theta_1 < \theta_0$

Then $\bar{\phi}(X) = 1 - \phi^*(X)$ (reject for __small__ T)

is LRT at level $\mathbb{E}_{\theta_0} \bar{\phi}(X) = 1 - \alpha$

$\Rightarrow$ $1 - \alpha \le \beta_{\bar{\phi}}(\theta_1) = 1 - \beta_{\phi^*}(\theta_1)$ for $\theta_1 < \theta_0$ $\boxtimes$

__Remark__ We also showed ϕ^* __minimizes__ $\mathbb{P}_\theta$(Type I error) for $\theta < \theta_0$ (among tests with $\mathbb{E}_{\theta_0} \phi(X) = \alpha$)