Chapter 8 hypothesis testing

Friday, May 31, 2024 08:39

Hypothesis: a statement about a population parameter

Hypothesis testing: based on the data (test stats), decide which of the two complementary hypothesis is true.

rejection region: The subset of the sample space for which H_0 will be rejected

acceptance region: The complement of the rejection region.

e.g. H_0: mean height is smaller than 160. ; test stat: $\bar{x}$

↳ reject H_0: if $\bar{x} > 160$

↳ Rejection region $\{(x_1, x_2 .. x_n) : \bar{x} > 160\}$

methods of finding tests

方法① Likelihood ratio test stats (related to MLE estimator)

$H_0: \theta \in \Theta_0$, $H_1: \theta \in \Theta_0^c$

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}$$ is the likelihood ratio test stat.

$\sup_{\Theta_0} L(\theta|\mathbf{x})$ → max prob of the observed sample over Θ_0

$\sup_{\Theta} L(\theta|\mathbf{x})$ → max prob of the observed samples over all possible params.

↳ reject H_0 if $\lambda(x) \le c$ $\quad c \in (0,1)$

↳ rejection region $\{\mathbf{x} : \lambda(\mathbf{x}) \le c\}$

relate to MLE

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

→ $\hat{\theta}_0$ is the MLE of θ over Θ_0. constrained/restricted

→ $\hat{\theta}$ is the MLE of θ over whole param space. unconstrained/unrestricted

Example

8.2.2 Setting: known variance

$x_1, x_2 \dots x_n$ iid $N(\theta, 1)$

$H_0: \theta = \theta_0$, $H_1: \theta \ne \theta_0$

$\sup_{\Theta_0} L(\theta|\mathbf{x}) = L(\theta_0|\mathbf{x})$ $\quad \Theta_0$: contains only one value θ_0

unrestricted $\sup_{\Theta} L(\theta|\mathbf{x}) = L(\hat{\theta}_{MLE}|\mathbf{x}) = L(\bar{x}|\mathbf{x})$

$$\therefore \lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\bar{x}|\mathbf{x})} = \exp\{-n(\bar{x}-\theta_0)^2/2\}$$

↳ reject H_0 if $\exp\{-n(\bar{x}-\theta_0)^2/2\} \le c$

↳ rejection region $\{\mathbf{x} : |\bar{x}-\theta_0| \ge \sqrt{-2\log c/n}\}$

→ reject when $\bar{x}$, θ_0 差别 a lot

setting: unknown variance

$x_1, x_2 \dots x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$H_0: \mu \le \mu_0$, $H_1: \mu > \mu_0$. σ^2 nuisance parameter

$$\lambda(\mathbf{x}) = \frac{\max_{\{\mu,\sigma^2 : \mu \le \mu_0, \sigma^2 > 0\}} L(\mu, \sigma^2|\mathbf{x})}{\max_{\{\mu, \sigma^2, \mu \in \mathbb{R}, \sigma^2 > 0\}} L(\mu, \sigma^2|\mathbf{x})} = \frac{L(\mu_0, \hat{\sigma}_0^2|\mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})}$$

$$\lambda(x) = \begin{cases} 1, & \text{if } \hat{\mu} \le \mu_0 \\ \dfrac{L(\mu_0, \hat{\sigma}_0|\mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})}, & \text{if } \hat{\mu} > \mu_0 \end{cases}$$

$\hat{\mu}$ μ_0

σ^2 is the nuisance param. $\hat{\sigma}_0^2 = \Sigma(x_i - \mu_0)^2/n$.

Example $x_1 \dots x_n$ exponential population

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x \ge \theta \\ 0, & x < \theta \end{cases}$$

$$L(\theta|\mathbf{x}) = \begin{cases} e^{-\Sigma(x_i - \theta)}, & \theta \le \min_i x_i \\ 0, & \text{otherwise} \end{cases}$$

$$L(\theta|x_{(1)}) = \begin{cases} ne^{-n(x_{(1)}-\theta)}, & \theta \le x_{(1)} \\ 0, & \theta > x_{(1)} \end{cases}$$

* not differentiable at $\theta = x_{(1)}$

cannot use score function to find MLE

$H_0: \theta \le \theta_0$, $H_1: \theta > \theta_0$.

$\sup_{\Theta} L(\theta|\mathbf{x}) = e^{-\Sigma x_i + n\theta} = e^{-\Sigma x_i + n \cdot x_{(1)}}$ $\quad L(x_{(1)}|\mathbf{x})$

→ increasing fun. $\hat{\theta}_{MLE} = \max \theta = \min_i x_i = x_{(1)}$

$\sup_{\Theta_0} L(\theta|\mathbf{x}) = e^{-\Sigma x_i + n\theta} \Rightarrow L(\theta_0|\mathbf{x})$

→ under H_0. $\max \theta$ is θ_0

① $x_{(1)}$ θ_0 ② θ_0 $x_{(1)}$

$$\therefore \lambda(\mathbf{x}) = \begin{cases} 1, & x_{(1)} \le \theta_0 \\ e^{n\theta_0 - nx_{(1)}}, & x_{(1)} > \theta_0 \end{cases}$$

↳ reject $\lambda(\mathbf{x})$ if $e^{n\theta_0 - nx_{(1)}} \le c$

↳ rejection region $x_{(1)} \ge \theta_0 - \frac{\log c}{n}$

$T(\mathbf{X})$ is a suffi statistic for θ

$\lambda^*(t)$, $\lambda(\mathbf{x})$ are LRT statistics based on T and $\mathbf{X}$

then, $\lambda^*(T(x)) = \lambda(\mathbf{x})$ for every $\mathbf{x}$

翻译为plain language 就是说求 $L(\theta|T(\mathbf{X}))$ 和 $L(\theta|\mathbf{X})$ 一样, 如果 $T(\mathbf{X})$ 是 suffi stat, 所以需要知道 $T(\mathbf{X})$ 的PDF

方法② Bayesian tests

- classical: hypothesis either true or false
  if $\theta \in \Theta_0$, $P(H_0 \text{ is true}|\mathbf{X}) = 1$ & $P(H_1 \text{ is true}|\mathbf{X}) = 0$
- Bayesian: if $P(\theta \in \Theta_0|\mathbf{X}) \ge P(\theta \in \Theta_0^c|\mathbf{X})$, accept H_0

  Rejection region: $\{\mathbf{X} : P(\theta \in \Theta_0^c|\mathbf{X}) > \frac{1}{2}\}$ → could use other threshold

  if $P(\theta \in \Theta_0|\mathbf{X}) < P(\theta \in \Theta_0^c|\mathbf{X})$, reject H_0

Method of evaluating tests

- when deciding to accept or reject null hypothesis H_0, an experimenter might make a mistake.
- evaluate hypothesis test through probability of making mistakes

error probabilities

type I: $\theta \in \Theta_0$, but hypothesis test reject H_0, reject region $\{\mathbf{X} : \mathbf{X} \in R\}$

type II: $\theta \in \Theta_0^c$, but hypothesis test accept H_0, accept region

ground truth.

For $\theta \in \Theta_0$, Type I error probability: $P_\theta(\mathbf{X} \in R)$

For $\theta \in \Theta_0^c$, Type II error probability: $P_\theta(\mathbf{X} \in R^c) = 1 - P_\theta(\mathbf{X} \in R)$

power function. of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.

$$\beta(\theta) = P_\theta(\mathbf{X} \in R) = \begin{cases} \text{probability of type I error}, & \text{if } \theta \in \Theta_0 \\ 1 - \text{probability of type II error}, & \text{if } \theta \in \Theta_0^c \end{cases}$$

都是 $P_\theta(\mathbf{X} \in R)$, 但一个是 Θ_0 下, 一个是 Θ_0^c 下.

Example $X \sim bin(5, \theta)$

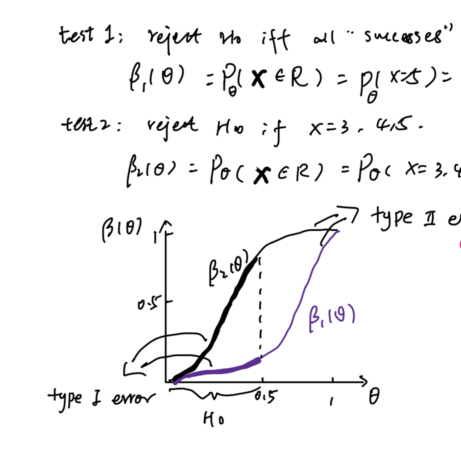
$H_0: \theta \le \frac{1}{2}$ $\quad H_1: \theta > \frac{1}{2}$

test 1: reject H_0 iff all "successes" are observed

$\beta_1(\theta) = P_\theta(\mathbf{X} \in R) = P_\theta(X=5) = \binom{5}{5}\theta^5(1-\theta)^0 = \theta^5$

test 2: reject H_0 if $x = 3, 4, 5$.

$\beta_2(\theta) = P_\theta(\mathbf{X} \in R) = P_\theta(X = 3, 4, \text{or } 5)$



Normal power function

$X_1, X_2 \dots X_n$ iid $N(\theta, \sigma^2)$, σ^2 known.

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

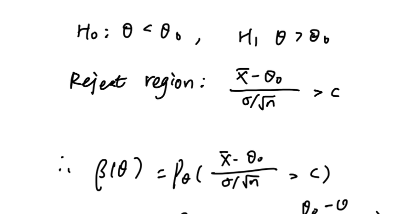
Reject region: $\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c$

$$\therefore \beta(\theta) = P_\theta\left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} > c\right) = P_\theta\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

→ $N(0,1)$

$$\frac{\bar{x} - \theta + \theta - \theta_0}{\sigma/\sqrt{n}} > c$$

$$\frac{\bar{x} - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}.$$



- wish Type I error < 0.1
  ∴ max $\beta(\theta)$ under H_0 < 0.1
  $\beta(\theta_0) < 0.1$
- whish Type II error of 0.2 if $\theta > \theta_0 + \sigma$
  $\beta(\theta_0 + \sigma) = 1 - \text{type II} = 1 - 0.2 = 0.8$

- if n fixed, cannot min both type I and type II error simultaneously
- restrict tests that control the type I error at a specified level, then search for tests that have type II error that is as small as possible.

① Size α test: $\alpha \in (0,1)$ a test with $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

② level α test: $\alpha \in (0,1)$ a test with $\sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha$

② contains ①.

\# Sometimes computationally impossible to construct a size α test (always exists)

Example: size of LRT

example 8.2.2 we have $R: \{\mathbf{x} : |\bar{x} - \theta_0| \ge \sqrt{-2\log c/n}\}$

α size LRT is

$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in R) = P_{\theta_0}(|\bar{X} - \theta_0| \ge \sqrt{-2\log c/n}) = \alpha$

[illegible]

Most powerful tests

Fix the type I error in the class C, a test in this class is uniformly most powerful (UMP) class C test if $\beta(\theta) \ge \beta'(\theta)$ for every $\theta \in \Theta_0^c$, $\beta'(\theta)$ is a power function of a test in class C.

翻译为plain language: [illegible] type I [illegible] $\beta(\theta)$ [illegible] type II min [illegible] for $\theta \in \Theta_0^c$.

Neyman-Pearson Lemma/test

$H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$ with pdf $f(\mathbf{x}|\theta_i)$ $i = 0, 1$.

using a test with rejection region R that satisfies

$\mathbf{x} \in R$ if $f(\mathbf{x}|\theta_1) > k f(\mathbf{x}|\theta_0)$ ①

and $\mathbf{x} \in R^c$ if $f(\mathbf{x}|\theta_1) < k f(\mathbf{x}|\theta_0)$

$P_{\theta_0}(\mathbf{X} \in R) = \alpha$ ②

is uniformly most powerful (UMP) if ① ② satisfied

also apply to suffi stats. with pdf $g(t|\theta_i)$, rejection region S.

$P_{\theta_0}(T \in S) = \alpha$

and $t \in S$ if $g(t|\theta_1) > k\, g(t|\theta_0)$

$t \in S^c$ if $g(t|\theta_1) < k\, g(t|\theta_0)$

解释 UMP 的test

① $\frac{f(x|\theta_1)}{f(x|\theta_0)} > k$

② $\frac{g(t|\theta_1)}{g(t|\theta_0)} > k$

③ this is the UMP

Karlin-Rubin

T is a suffi. stats for θ

$\{g(t|\theta) : \theta \in \Theta\}$ of T has MLR. → t的PDF是关于θ的单调函数.

for any t_0, the test that rejects H_0 iff $T > t_0$, is UMP- level α test.

↳ a fun of μ_0 (see example)

where $\alpha = P_{\theta_0}(T > t_0)$

①首先找到 T 为 suffi

②然后判断 T的PDF是关于θ的单调函数

③ 找到 $P(T > c|H_0) = \alpha$ 中 c的取值, 则

④ $T > c$ (取值) 是一个UMP.

$H_0: \mu = \mu_0$, $H_1: \mu > \mu_0$

$T = \bar{X}$

$P(\bar{X} > c|H_0) = \alpha$

under H_0,

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

$$P\left(Z > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha = 0.05$$

a R.V. x with param θ has monotone likelihood ratio if

for every $\theta_2 > \theta_1$, $f(x|\theta_2)/f(x|\theta_1)$ is a monotone (nonincreasing/nondecreasing) function of x on $\{x : f(x|\theta_1) > 0 \text{ or } f(x|\theta_2) > 0\}$.

exponential family with $f(x|\theta) = h(x)c(\theta)e^{w(\theta)x}$ has MLR if $w(\theta)$ is a nondecreasing function.

① $T = \bar{X}$ $\bar{X}$ 是 suffi

② $\bar{X} = \frac{1}{n}\Sigma x_i$

③ 则 $P(\bar{X} > c) = \alpha$

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha = 0.05$$

$$Z > \frac{c - \mu_0}{\sigma/\sqrt{n}} = Z_\alpha \text{ (单侧)}$$

$\Rightarrow c = \mu_0 + Z_\alpha \cdot \sigma/\sqrt{n}$

UMP test为 $\bar{X} > \mu_0 + Z_\alpha \cdot \sigma/\sqrt{n}$

A p-value $p(\mathbf{X})$ is a test stat that satisfies $0 \le p(\mathbf{X}) \le 1$ for every sample $\mathbf{x}$. A p-value is valid if, for every $\theta \in \Theta_0$ and $\alpha \in (0,1)$,

$$P_{\theta_0}(p(\mathbf{X}) \le \alpha) \le \alpha$$

a level α test is $P_{\theta_0}(\mathbf{X} \in R) \le \alpha$. so;

reject H_0 iff $p(\mathbf{X}) \le \alpha$ is a level α test

→ rejection region.

$$p(\mathbf{X}) = \sup_{\theta \in \Theta_0} P_\theta(W(\mathbf{X}) \ge W(\mathbf{x}))$$

→ observed value.

→ a test stat

$P(X \in R) = 0.049$

P60: small-sample inference ordinary p-value conservative.

conservative ⇒ ordinary p-value larger than they should be.

$P(W(X) \ge W(x))$

$P(Z \ge \text{sth}) \le \alpha$

size α test $X \ge 7.5$

但只能达到 $X = 8$,

则是一个 level-α test

① reject if $X > 7.1$

② discrete reject if $X > 8$

③