Asymptotic testing: basics and relative efficiencies

John Duchi

Stats 300b – Winter Quarter 2021

Outline

- Power and level of tests
- Sequences of local alternatives
- Comparison of tests

Reading: This previews some of what comes after, but

- van der Vaart, *Asymptotic Statistics* Ch. 14
- Lehmann & Romano, *Testing Statistical Hypothesis* Ch. 13.1, 13.2

Asymptotic level of a test

- Parameter θ of interest in family $\{P_\theta\}_{\theta\in\Theta}$
- Testing null $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$

Definition (Power function)

Given a sequences of test statistics T_n and critical regions K_n, (test rejects H_0 if $T_n \in K_n$), the *power function* is

$$\pi_n(\theta) := P_\theta(T_n \in K_n)$$

Definition

The *uniform asymptotic* and *pointwise asymptotic* levels of T_n for null H_0 are

$$\limsup_{n\to\infty} \sup_{\theta_0\in\Theta_0} \pi_n(\theta_0) \quad \text{and} \quad \sup_{\theta_0\in\Theta_0} \limsup_{n\to\infty} \pi_n(\theta_0)$$

How should we compare tests?

Idea 1: compare all powers

- Let tests $T_n^{(i)}$ have powers $\pi_n^{(i)}$. Then $T_n^{(1)}$ is *uniformly more powerful* than $T_n^{(2)}$ for testing $H_0 : \theta_0 \in \Theta_0$ against $H_1 : \theta_1 \in \Theta_1$ if

$$\pi_n^{(1)}(\theta) \leq \pi_n^{(2)}(\theta) \text{ for all } \theta \in \Theta_0$$
$$\pi_n^{(1)}(\theta) \geq \pi_n^{(2)}(\theta) \text{ for all } \theta \in \Theta_1$$

- unfortunately, way too strong

Idea 2: look at asymptotic power and level?

- unfortunately, all reasonable tests have asymptotic power 1.

Example: the sign test for location

- $X_i \overset{\text{iid}}{\sim} P(\cdot - \theta)$, P has symmetric density, $\theta \in \mathbb{R}$
- sign test of $H_0 : \theta = 0$ against $H_1 : \theta > 0$:

$$S_n := \frac{1}{n} \sum_{i=1}^{n} \text{sign}(X_i), \quad \text{so} \ \ \text{sign}(X_i) \overset{\text{iid}}{\sim} \text{Uni}\{\pm 1\} \text{ under } H_0$$

  while $\mu(\theta) := \mathbb{E}_\theta[S_n]$ satisfies $\mu(\theta) > 0$ under H_1
- reject H_0 if $\sqrt{n} S_n \geq z_{1-\alpha}$, where $\Phi(z_{1-\alpha}) = 1 - \alpha$ for standard normal CDF

Observation

$$\lim_{n \to \infty} \pi_n(\theta) = \begin{cases} \alpha & \textit{if } \theta = 0 \\ 1 & \textit{if } \theta > 0 \end{cases}$$

Large deviations?

Idea 3: let's use large deviations and information theory

- developed by Hoeffding and Chernoff
- study limits of
$$\frac{1}{n} \log \pi_n(\theta)$$
- if cumulant generating function $\varphi(\lambda) = \log(\mathbb{E}[e^{\lambda X}])$ exists,
$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq t) = \inf_{\lambda \geq 0} \{\varphi(\lambda) - \lambda t\}$$
- issue: doesn't readily generalize to estimation

Local alternatives

Idea: study problems getting *closer* to one another as $n \to \infty$, so H_0, H_1 are harder to distinguish

- *local perturbation* of

$$H_0 : \theta = \theta_0 \;\; \text{to} \;\; H_1 : \theta = \theta_0 + \frac{h}{\sqrt{n}}$$

where h is fixed will give "right" behavior

Example (Gaussian mean shifts)

Let $H_0 : X_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $H_1 : X_i \stackrel{\text{iid}}{\sim} \mathcal{N}(h/\sqrt{n}, 1)$

$$T_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \stackrel{\text{dist}}{=} \begin{cases} \mathcal{N}(0,1) & \text{under } H_0 \\ \mathcal{N}(h,1) & \text{under } H_1 \end{cases}$$

General idea

suppose there exists increasing mean function μ and variance σ s.t.

$$\sqrt{n}\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)} \xrightarrow[\theta_n]{d} \mathcal{N}(0,1) \quad \text{where} \quad \theta = \frac{h}{\sqrt{n}}$$

- then $\sqrt{n}(T_n - \mu(0)) \xrightarrow{d} \mathcal{N}(0, \sigma^2(0))$ for $\theta_n = 0$
- asymptotic level α test of $H_1 : \theta = 0$ against $H_1 : \theta > 0$

$$\text{reject if} \quad \sqrt{n}(T_n - \mu(0)) \geq \sigma(0) z_{1-\alpha}$$

Theorem

Assume $\mu'(0)$ exists and σ is continuous at 0. Then

$$\pi_n(\theta_n) \to 1 - \Phi\left(z_{1-\alpha} - h\frac{\mu'(0)}{\sigma(0)}\right) = \Phi\left(z_\alpha + h\frac{\mu'(0)}{\sigma(0)}\right)$$

Proof of theorem

Example: exponential families

exponential family model with density

$$p_\theta(x) = \exp(\theta^T x - A(\theta))$$

Observation

If $\theta_n \to \theta_0 \in \operatorname{int}\operatorname{dom} A$, *then for* $X_i^n \overset{\text{iid}}{\sim} P_{\theta_n}$ *and* $\mu_n = \mathbb{E}[X_i^n] = \nabla A(\theta_n)$,

$$\sqrt{n}(\overline{X}_n^n - \mu_n) \xrightarrow{d} \mathcal{N}(0, \nabla^2 A(\theta_0))$$

Proposition

For $\widehat{\theta}_n = \operatorname{argmin}_\theta\{-P_n \log p_\theta(X)\} = (\nabla A)^{-1}(\overline{X}_n^n)$,

$$\sqrt{n}(\widehat{\theta}_n - \theta_n) \xrightarrow[\theta_n]{d} \mathcal{N}(0, \nabla^2 A(\theta_0)^{-1})$$

Slope of a test

Definition

The *slope* of a sequence of tests T_n is $\mu'(0)/\sigma(0)$

idea: if slope is big, test is powerful:

$$\begin{aligned} 1 - \Phi\left(z_{1-\alpha} - h\frac{\mu'(0)}{\sigma(0)}\right) &= \Phi\left(z_\alpha + h\frac{\mu'(0)}{\sigma(0)}\right) \\ &= \alpha + h\frac{\mu'(0)}{\sigma(0)}\phi(z_\alpha) + O(h^2) \end{aligned}$$

Relative efficiency of tests

- indices $\nu \in \mathbb{N}$, $\nu \to \infty$
- tests $H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_\nu$
- for level α and power β, define

$$n_\nu := n_\nu(\alpha, \beta) = \inf \left\{ n \in \mathbb{N} : \pi_n(0) \le \alpha,\ \pi_n(\theta_\nu) \ge \beta \right\},$$

  smallest number of observations to distinguish $\theta = 0$ from $\theta = \theta_\nu$

Definition (Asymptotic relative efficiency / Pitman efficiency)

For tests $T_n^{(1)}$ and $T_n^{(2)}$ with distinguishing numbers $n_\nu^{(i)}$, the *asymptotic relative efficiency* of $T^{(1)}$ w.r.t. $T^{(2)}$ is

$$\lim_{\nu \to \infty} \frac{n_\nu^{(2)}}{n_\nu^{(1)}}.$$

An exact calculation of error

Definition

The *total variation distance* between distributions P and Q is

$$\|P - Q\|_{\mathsf{TV}} := \sup_A |P(A) - Q(A)| = \frac{1}{2}\int |p - q| d\mu$$

Lemma (Le Cam)

The optimal test $\Psi : \mathcal{X} \to \{0, 1\}$ *of* P *against* Q *satisfies*

$$\inf_\Psi \{P(\Psi \neq 0) + Q(\Psi \neq 1)\} = 1 - \|P - Q\|_{\mathsf{TV}}$$

Asymptotic relative efficiency via slopes

Theorem (14.19 in van der Vaart, 13.2.1 in TSH)

Let models $\{P_{n,\theta}\}$ *satisfies* $\lim_{\theta \to 0} \|P_{n,\theta} - P_{n,0}\|_{\mathsf{TV}} = 0$. *Assume*

$$\sqrt{n}\frac{T_n^{(i)} - \mu_i(\theta_n)}{\sigma_i(\theta_n)} \xrightarrow{d} \mathcal{N}(0,1)$$

when $\theta_n \to 0$ *and* σ_i *is continuous with* $\mu_i'(0) > 0$. *Then the ARE of* $T^{(1)}$ *against* $T^{(2)}$, *rejecting* $H_0 : \theta = 0$ *when* T_n *is large, is*

$$\left(\frac{\mu_1'(0)/\sigma_1(0)}{\mu_2'(0)/\sigma_2(0)}\right)^2 \quad \textit{for any } \theta_n \to 0, \alpha < \beta$$

Location tests, revisited

Symmetric distribution with cdf F, $X_i \stackrel{\text{iid}}{\sim} F(\cdot - \theta_n)$ for $\theta_n > 0$

Example (The sign test)

For test rejecting if $\sqrt{n} S_n \ge z_{1-\alpha}$,

$$\mu(\theta) = \mathbb{E}_\theta[S_n] = 2F(\theta) - 1, \quad \sigma^2(\theta) = 1 - (2F(\theta) - 1)^2$$

$\mu'(0) = 2f(0)$ and $\sigma^2(0) = 1$. Asymptotic power for $\theta_n = h/\sqrt{n}$:

$$\pi_n(\theta_n) \to \Phi(z_\alpha + 2hf(0))$$

Example (T-tests)

Rejects if $\overline{X}_n / \widehat{\sigma}_n$ large, $\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2$. Asymptotics:

$$\sqrt{n} \left(\frac{\overline{X}_n}{\widehat{\sigma}_n} - \frac{h/\sqrt{n}}{\sigma} \right) \xrightarrow[h/\sqrt{n}]{d} \mathcal{N}(0, 1)$$

and $\sigma^2(\theta) = 1$, $\mu(\theta) = \frac{\theta}{\sigma}$

Comparing the sign and T-test

Symmetric density $x \mapsto f(x-\theta)$, testing $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$

Slope of sign:	slope of T:
$2f(0)$	$\mathrm{Var}_f(X)^{-1/2} = \frac{1}{\sqrt{\int x^2 f(x) dx}}$

some cases:

- standard normal: slopes $\sqrt{\frac{2}{\pi}}$ versus 1, so T-test has relative efficiency $\pi/2 \approx 1.57$
- Laplace: $f(x) = \frac{1}{2}e^{-|x|}$, slopes 1 versus $\frac{1}{\sqrt{2}}$ so sign test has relative efficiency 2

rough takeaway: fatter tails make the T-test worse and sign test more "robust"