Stats 300b: Theory of Statistics	**Winter 2017**

Lecture 17 – March 7

Lecturer: John Duchi *Scribe: Nikolaos Ignatiadis, Michael Sklar*

Warning: *these notes may contain factual errors*

Reading: VDV (van der Vaart, Asymptotic Statistics) Chapter 14: Relative efficiency of tests

Outline: In this lecture we broadly cover *Asymptotic testing*. In particular, we cover the following:

- Asymptotic power and level of tests
- Sequence of local alternatives
- Comparison of tests

Basic setup

Consider the basic problem of testing the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1$.

Given a test statistic T_n and critical region K_n, we reject the null hypothesis H_0 if $T_n \in K_n$. The power function of T_n (and the corresponding test) based on rejection region K_n then is:

$$\pi_n(\theta) = \mathbb{P}_\theta[T_n \in K_n]$$

Definition 1. The sequence of tests based on statistics T_n and rejection regions K_n is asymptotically level (size) α if:

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_0} \mathbb{P}_\theta[T_n \in K_n] \le \alpha$$

Naïve comparison of tests

We are interested now in answering the following question: How should we compare two tests of similar levels?

To make this precise, let $\pi_n^{(1)}, \pi_n^{(2)}$, be power functions for tests $(T_n^{(1)}, K_n^{(1)}), (T_n^{(2)}, K_n^{(2)})$ and we want to compare the power of the tests.

Attempt 1: An uncontroversial way to say that test 1 is better than test 2 is if the following two conditions are satisfied:

1. $\pi_n^{(1)}(\theta) \le \pi_n^{(2)}(\theta) \quad \forall \quad \theta \in \Theta_0$
2. $\pi_n^{(1)}(\theta) \ge \pi_n^{(2)}(\theta) \quad \forall \quad \theta \in \Theta_1$

In fact, if also $\pi_n^{(1)}(\theta) > \pi_n^{(2)}(\theta)$ for some $\theta \in \Theta_1$, then test 1 would uniformly dominate test 2 (and test 2 would be inadmissible).

However, this is an extremely strong requirement and will not happen, except in extremely simple cases (such as simple hypotheses for which the Neyman-Pearson Lemma applies).

Attempt 2: Instead, let's try to take limits as $n \to \infty$. Unfortunately this does not work either:

Example 1 (Sign test for location)**:** Let $X_i \sim \mathbb{P}[\cdot - \theta]$ i.i.d., where $\mathbb{P}$ is symmetric and has density at $\theta = 0$. Let $H_0 : \theta = 0$ and $H_1 : \theta > 0$. Consider the sign test based on the statistic:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} \operatorname{sign}(X_i)$$

Under H_0, $\operatorname{sign}(X_i)$ are just i.i.d. Rademacher, i.e. random signs (Uniform on {-1,+1}). Now let:

$$\mu(\theta) = \mathbb{E}_\theta[\operatorname{sign}(X_i)] = \mathbb{E}_0[\operatorname{sign}(X_i + \theta)]$$

By the above comment, we clearly have that $\mu(0) = 0$. On the other hand, for $\theta > 0$, we have that $\mu(\theta) > 0$, since $\mathbb{P}_0[X_i + \theta > 0] = \mathbb{P}[X_i > -\theta] > \mathbb{P}[X_i > 0] = \frac{1}{2}$.

Furthermore since $\operatorname{Var}_0(\operatorname{sign}(X_i)) = 1$, we have by the Central Limit Theorem that:

$$\text{Under } H_0 : \quad \sqrt{n} S_n \xrightarrow[H_0]{\mathcal{D}} \mathcal{N}(0, 1)$$

Now the natural (asymptotic) level α test rejects if $\sqrt{n} S_n \geq z_\alpha$, where z_α is the $1 - \alpha$ quantile of the Standard Normal distribution, i.e. $\mathbb{P}[\mathcal{N}(0, 1) \geq z_\alpha] = \alpha$.

We next turn to study the power function $\pi_n(\theta)$ for $\theta > 0$: What happens asymptotically?

$$\pi_n(\theta) = \mathbb{P}_\theta[\sqrt{n} S_n \geq z_\alpha] = \mathbb{P}_\theta[\sqrt{n}(S_n - \mu(\theta)) \geq z_\alpha - \sqrt{n}\mu(\theta)]$$

Now observe that, again by the CLT:

$$\sqrt{n}(S_n - \mu(\theta)) \xrightarrow[\theta]{\mathcal{D}} \mathcal{N}(0, \sigma^2(\theta))$$

where $\sigma^2(\theta) = \operatorname{Var}_\theta(\operatorname{sign}(X_i))$ and furthermore since $\theta > 0$ ($\Rightarrow \mu(\theta) > 0$) we have that

$$z_\alpha - \sqrt{n}\mu(\theta) \to -\infty \text{ as } n \to \infty$$

Thus:

$$\pi_n(\theta) \xrightarrow[\theta > 0]{n \to \infty} \mathbb{P}[\sigma(\theta) Z \geq -\infty] = 1$$

where $Z \sim \mathcal{N}(0, 1)$. ♣

Take-home message: Any sensible level α test will be consistent against all alternatives and have $\lim_{n \to \infty} \pi_n(\theta) = 1$ for all $\theta \in \Theta_1$. Hence we cannot really compare reasonable tests using this type of asymptotic analysis.

More fine-grained asymptotic comparison of tests

So the key question is: What should we do to understand limits and power of tests? Here there are two main key ideas that provide an answer:

Idea 1 (Hoeffings, '60s): Use large deviations and information theory by studying (for $\theta \in \Theta_1$):

$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \pi_n(\theta))$$

The intuition here is that often $\pi_n(\theta) \to 1$ exponentially fast in $n \to \infty$. For example, in the Gaussian case we approximately have that:

$$\pi_n(\theta) \asymp 1 - \exp\left(-\frac{n\theta^2}{2}\right)$$

Idea 2 (Le Cam, '70s): The idea here is to look at sequences of problems getting "closer" together or harder to distinguish as $n \to \infty$.

We consider the following thought experiment: Let $H_0 : \theta = 0$ and $H_1 : \theta = \theta_n = \frac{h}{\sqrt{n}}$, where h is some fixed vector. This will give the "right" behaviour for "normal" models:

Example 2: Let $X_i \sim \mathcal{N}(0, 1)$ i.i.d. under the null hypothesis and $X_i \sim \mathcal{N}\left(\frac{h}{\sqrt{n}}, 1\right)$ i.i.d. under the alternative for $i = 1, \ldots, n$. Also consider the statistic:

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

We have:

$$\text{Under } H_0 : \quad T_n \sim \mathcal{N}(0, 1)$$
$$\text{Under } H_1 : \quad T_n \sim \mathcal{N}(h, 1)$$

To see this for the alternative, just observe that $T_n = h + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i$ with $W_i \sim \mathcal{N}(0, 1)$ i.i.d. This trivially also implies that asymptotically:

$$T_n \xrightarrow[H_0]{\mathcal{D}} \mathcal{N}(0, 1)$$

$$T_n \xrightarrow[\frac{h}{\sqrt{n}}]{\mathcal{D}} \mathcal{N}(h, 1)$$

♣

The idea now is that any limiting test must distinguish between $\mathcal{N}(0, 1)$ and $\mathcal{N}(h, 1)$. Here we started with a Gaussian example, but this limiting normality will end up holding under extraordinary generality in many problems. This will allow us to reduce problems asymptotically to testing or estimation in Gaussian location models.

In particular, now suppose that there exists a mean function $\mu(\theta)$ and a variance function $\sigma^2(\theta)$ such that:

$$\sqrt{n}\left(\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)}\right) \xrightarrow[\theta_n]{\mathcal{D}} \mathcal{N}(0, 1)$$

Here we let $\theta_n = \frac{h}{\sqrt{n}}$. Also we will assume for simplicity that μ is differentiable in θ with $\mu'(\theta) > 0$ for all $\theta \geq 0$.

The natural test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ is to reject if $\sqrt{n}\left(\frac{T_n - \mu(0)}{\sigma(0)}\right)$ is large, i.e. if:

$$\sqrt{n}\left(\frac{T_n - \mu(0)}{\sigma(0)}\right) \geq z_\alpha$$

By construction this is definitely asymptotically level α, since:

$$\limsup_{n\to\infty} \pi_n(0) = 1 - \Phi(z_\alpha) = \alpha$$

What about the limiting power under θ_n?

$$\pi_n(\theta_n) = \mathbb{P}_{\theta_n}[\sqrt{n}(T_n - \mu(\theta_n)) \geq \sigma(0)z_\alpha - \sqrt{n}(\mu(\theta_n) - \mu(0))]$$

What is the limit of $\sqrt{n}(\mu(\theta_n) - \mu(0))$? By differentiability of μ it is just:

$$\sqrt{n}(\mu(\theta_n) - \mu(0)) = h\frac{\mu\left(\frac{h}{\sqrt{n}}\right) - \mu(0)}{\frac{h}{\sqrt{n}}} \xrightarrow{n\to\infty} h\mu'(0)$$

Recalling that $\sqrt{n}\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)} \xrightarrow[\theta_n]{\mathcal{D}} \mathcal{N}(0,1)$, if we additionally assume that $\sigma(\theta_n) \to \sigma(0)$ as $\theta_n \to 0$, we get the following theorem:

Theorem 1. *If $\mu'(0)$ exists, $\sigma\left(\frac{h}{\sqrt{n}}\right) \to \sigma(0)$ as $n \to \infty$, then the level α test rejecting large values of $\sqrt{n}(T_n - \mu(0))$ satisfies:*

$$\pi_n\left(\frac{h}{\sqrt{n}}\right) \xrightarrow{n\to\infty} 1 - \Phi\left(z_\alpha - h\frac{\mu'(0)}{\sigma(0)}\right)$$

Here Φ is the standard normal distribution function.

The intuition is the following: If $\mu'(0) \gg 0$, then the test has good power, since

$$\Phi\left(z_\alpha - h\frac{\mu'(0)}{\sigma(0)}\right) \approx 0$$

.

Definition 2. If

$$\sqrt{n}\left(\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)}\right) \xrightarrow[\theta_n]{\mathcal{D}} \mathcal{N}(0,1)$$

where $\theta_n = \frac{h}{\sqrt{n}}$, then the slope of the tests T_n is defined as $\frac{\mu'(0)}{\sigma(0)}$.

We will use this slope to compare tests to one another.

Relative Efficiency of Tests

Consider a sequence of simple tests: For $\nu \in \mathbb{N}$, let $H_0 : \theta = 0$ and $H_1 : \theta = \theta_\nu$, where $\theta_\nu \to 0$ as $\nu \to \infty$. Fix a level α and power $\beta \in (\alpha, 1)$. Define the distinguishing number $n_\nu := \inf\{n \in \mathbb{N} : \pi_n(0) \leq \alpha, \pi_n(\theta_\nu) \geq \beta\}$. In other words, n_ν is the smallest number of observations necessary to distinguish H_0 from H_1 at level α and power β. Let tests $T_n^{(1)}, T_n^{(2)}$ have distinguishing numbers $n_\nu^{(1)}, n_\nu^{(2)}$, respectively.

Definition 3. The Asymptotic Relative Efficiency (ARE), or Pitman Efficiency, of $T^{(1)}$ relative to $T^{(2)}$ is

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

So, an ARE of 2 means that $T^{(2)}$ asymptotically requires twice the sample size of $T^{(1)}$ to get the same power and level. This definition implicitly assumes that the ARE is independent of α and β. That assumption will be true for the next theorem, and is often true in general.

Definition 4. The total variation distance between two probability distributions P and Q is

$$\|P - Q\|_{TV} = \sup_A |P(A) - Q(A)|$$

Theorem 2. *(van der Vaart 14.19) Let models $\{P_{n,\theta}\}_{\theta \geq 0}$ satisfy*

$$\lim_{\theta \to 0} \|P_{n,\theta} - Q_{n,\theta}\|_{TV} = 0$$

Let tests $T^{(1)}$, $T^{(2)}$ satisfy that as $\theta_n \downarrow 0$,

$$\sqrt{n}\left(\frac{T_n^{(i)} - \mu_i(\theta_n)}{\sigma_i(\theta_n)}\right) \xrightarrow[\theta_n]{\mathcal{D}} \mathcal{N}(0, 1)$$

where $i \in \{1, 2\}$, σ_i is continuous at 0, and $\mu_i'(0) > 0$. Then the ARE of tests rejecting $H_0 : \theta = 0$ against $H_1 : \theta > 0$ when $T_n^{(i)}$ is large is

$$\left(\frac{\mu_1'(0)/\sigma_1(0)}{\mu_2'(0)/\sigma_2(0)}\right)^2$$

for any α, β such that $0 < \alpha < \beta < 1$.

Example 3 (Sign test)**:** $H_0 : \theta = 0; H_1 : \theta > 0$. Consider a location model where the CDF of x under θ is $F(\cdot - \theta)$, and Median(F) = 0.

We reparametrize the previous notation from Example 1, defining

$$S_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \geq 0\}}$$

As we found earlier,

$$\text{Under } H_0: \quad \sqrt{n}\left(S_n - \frac{1}{2}\right) \xrightarrow[H_0]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}\right)$$

With rejection of H_0 when

$$\sqrt{n}\big(S_n - 1/2\big) \geq z_\alpha/2$$

.

Then,

$$\mathbb{E}_{\theta_n}[S_n] = 1 - F(-\theta_n)$$

$$\sigma^2(\theta_n) = F(-\theta_n)(1 - F(-\theta_n))$$

$$\pi_n(\theta_n) = 1 - \Phi\left(\frac{z_\alpha/2 - \sqrt{n}(F(0) - F(-\theta_n))}{\sigma(\theta_n)}\right) + o(1)$$

If $F'(0) = f(0) > 0$ and $\theta_n \asymp \frac{1}{\sqrt{n}}$, then

$$\sqrt{n}(F(0) - F(-\theta_n)) = \sqrt{n}\theta_n f(0) + o(\sqrt{n}\theta_n)$$

So,

$$\pi_n\left(\frac{h}{\sqrt{n}}\right) \to 1 - \Phi(z_\alpha - 2hf(0))$$

Thus, the sign test performs well when the density at $\theta = 0$ is $\gg 0$. ♣

Example 4 (T-test)**:** In the same location-family setting, we reject when $\bar{X}/s$ is large, where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$s = \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}$$

By Slutsky's Theorem,

$$\sqrt{n}\left(\frac{\bar{X}}{s} - \frac{h/\sqrt{n}}{\sigma}\right) \xrightarrow[h/\sqrt{n}]{\mathcal{D}} \mathcal{N}(0,1)$$

If the location family has variance σ^2, then we satisfy Theorem 2's conditions with $\sigma^2(\theta_n) = 1$ and $\mu(\theta) = \theta/\sigma$. Thus, $\mu'(\theta) = 1/\sigma$.

Hence by the theorem, the corresponding slope will be

$$\frac{\mu'(0)}{\sigma(0)} = \frac{1/\sigma}{1} = \frac{1}{\sqrt{Var_0(X)}}$$

Comparing against the sign test for location models with symmetric densities, $H_0 : \theta = 0$, $H_1 : \theta > 0$, the slope of the sign is $2f(0)$.

For a standard Gaussian: the t-test slope is 1 and the sign test slope is $\sqrt{2/\pi}$, so the t-test is asymptotically $\pi/2$ times more efficient.

For a Laplace density:

$$f(x) = \frac{e^{-|x|}}{2}$$

The slopes are $1/\sqrt{2}$ for the t-test and 1 for the sign test. Thus, the ARE of a t-test against the sign test is $1/2$.

The sign test often outperforms the t-test when the tails are slightly fatter than Gaussian. ♣