Stat210B: Theoretical Statistics **Lecture Date: March 22, 2007**

Lecture 19: Asymptotic Relative Efficiency

Lecturer: Michael I. Jordan *Scribe: Kurt Miller*

1 Asymptotic Relative Efficiency

We will be working in the context of hypothesis testing where we have

$$\begin{aligned} \text{null hypothesis } H_0: &\quad \theta \in \Theta_0 \\ \text{alternative } H_1: &\quad \theta \in \Theta_1 \end{aligned}$$

and $\Theta_0 \cup \Theta_1$ is typically exhaustive and non-overlapping.

We will denote the *critical region* by K_n and the *power function* is $\pi_n(\theta) = P_\theta(T_n \in K_n)$ where T_n is our test statistic. The *size of the test* is $\sup\{\pi_n(\theta) : \theta \in \Theta_0\}$. A test is of level α if its size is $\leq \alpha$.

A test is *asymptotically* of size α if

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_0} \pi_n(\theta) \leq \alpha$$

We will look at the limiting power function of shrinking alternatives

$$\pi(h) = \lim_n \pi_n(h/\sqrt{n})$$

Theorem 1. *Theorem 14.7 in (**?**, p. 195). Assume without loss of generality that $\theta_0 = 0$ and therefore that our shrinking set of alternatives is $\theta_n = h/\sqrt{n}$. Also assume that*

$$\sqrt{n}\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)} \overset{\theta_n}{\rightsquigarrow} \mathcal{N}(0,1)$$

for some μ and σ where μ is differentiable at 0 and σ is continuous at 0. Then the tests that reject for large values of T_n and are asymptotically of level α have power functions that satisfy

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

where $\mu'(0)/\sigma(0)$ is called the slope (or efficiency) of the test.

Proof.

$$\begin{aligned} \pi_n(h/\sqrt{n}) &\triangleq P_{\theta_n}\left(\sqrt{n}(T_n - \mu(0)) > z_\alpha \sigma(0)\right) \\ &= P_{\theta_n}\left(\sqrt{n}(T_n - \mu(h/\sqrt{n})) > z_\alpha\sigma(0) - \sqrt{n}(\mu(h/\sqrt{n}) - \mu(0))\right) \\ &\to 1 - \Phi(z_\alpha - h\mu'(0)/\sigma(0)) \end{aligned}$$

where we have arrived at the last line since

$$\frac{\mu(h/\sqrt{n}) - \mu(0)}{h/\sqrt{n}} \to \mu'(0)$$

and by continuity of σ, we have $\sigma(\theta_n) \to \sigma(0)$. □

Definition 2 (Asymptotic relative efficiency)**.** The *Asymptotic Relative Efficiency* (ARE) is the ratio of the squares of slopes between two statistics.

Example 3 (Sign test)**.** This example is from Van der Vaart, but presents a different derivation than is found in the book.

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. from a symmetric density $f(x-\theta)$.
- Let our test statistic be $S_n = \frac{1}{n}\sum_i 1_{X_i>0}$
- At $\theta = 0$, this is the sum of Bernoulli variables with probability 1/2, so $\sigma^2(0) = 1/4$.

Remember that location families are qmd (as mentioned in lecture on 3/8) and that for qmd families, $P_{\theta+h/\sqrt{n}} \lhd P_\theta$ (see example 14 from the 3/13 lecture notes). Also, for location families, $\dot{\ell}_0 = -f'(x)/f(x)$. By contiguity, to get the distribution of the statistics under P_{θ_n}, compute τ_{12}. To compute τ_{12}, we note that this statistic is an (asymptotically) linear statistic, so by the results from example 4 from the 3/15 lecture notes), τ_{12} is equal to

$$\begin{aligned}
\tau_{12} &= \mathrm{Cov}_0(1_{X>0}, h^\top \dot{\ell}_0(X)) \\
&= -h \int 1_{x>0} \frac{f'(x)}{f(x)} f(x)dx \\
&= -h \int_0^\infty f'(x)dx \\
&= hf(0)
\end{aligned}$$

Therefore the slope of the sign test is $2f(0)$. It is good to have larger slopes, so this test will be better if there is a lot of mass around the origin.

Example 4 (t-test)**.** In example 5 from the 3/15 lecture notes, we showed that for the t-statistic, $\tau_{12} = h/\sigma$ and the variance under the null is 1, so the slope of the t-statistic is

$$\frac{1}{(\int x^2 f(x)dx)^{1/2}}$$

Therefore, we compute that the ARE of the sign test vs. the t-test is $4f^2(0)\int x^2 f(x)dx$. If this is > 1, then the sign test is better, if it is < 1, the t-test is better. We compute the AREs as

f distribution	ARE(sign, t)	t-test better?
logistic	$\pi^2/12$	yes
normal	$2/\pi$	yes
Laplace	2	no
uniform	1/3	yes

Note: Why do we square the slopes in the ARE ratio? This leads to ARE(A,B) being the number of points for A to be competitive with B. We will return to this later.

It turns out that above, we can show that 1/3 is a lower bound on the ARE using calculus of variations. Therefore, the uniform is the worst case for the sign test compared to the t-test.

Example 5 (2 sample tests for shift of location of $f(x-\theta)$ under the null $\theta = 0$ vs. $\theta > 0$)**.** The Mann-Whitney test for m and n samples respectively of x_i and y_j rejects if

$$\frac{1}{mn}\sum_i \sum_j 1_{x_i \le y_j}$$

is large. We will compare this to the two sample t-test. The Mann-Whitney statistic is a U-statistic, so we have a formula for computing its variance (or use contiguity) to get that the slope is $\int f dF/\sigma(0)$. Looking up the slope of the two-sample t-test, the ARE between the Mann-Whitney test and the t-test is $12 \text{Var} X (\int f^2(y)dy)^2$. This gives the chart

f distribution	ARE(M-W, t)	t-test better?
logistic	$\pi^2/9$	no
normal	$3/\pi$	yes
Laplace	$3/2$	no
uniform	1	equivalent
$t_{3-\text{dof}}$	1.24	no
$t_{3-\text{dof}}$	1.9	no
$c(1-x^2) \vee 0$	108/125	yes

Again, it can be shown using calculus of variations that 108/125 is the worst that the ARE can be, so the Mann-Whitney test can never be that much worse than the t-test, but it can be much better. The above two tables support our intuition that the sign test can lose a decent amount of information since it throws out a lot of information, but rank tests don't throw out as much, so they are competitive.

2 Interpreting the ARE

As mentioned above, the ARE is the number of points needed to achieve the same power between two tests. Let θ_ν be an alternative sequence where $\nu \to \infty$ where ν is the rate of gathering data. Define n_ν to be the minimal number of observations such that $\pi_{n_\nu}(0) \leq \alpha$ (since $\theta = 0$ is the null and the power at the null = the type-1 error rate, which we want to be $\leq \alpha$) and $\pi_{n_\nu}(\theta_\nu) \geq \gamma$ for some $\gamma \in (\alpha, 1)$. Do this for two tests, yielding $n_{\nu,1}$ and $n_{\nu,2}$ for tests 1 and 2, respectively.

Definition 6 (Pitman efficiency)**.** The *Pitman efficiency* is defined as

$$\lim_{\nu \to \infty} \frac{n_{\nu,2}}{n_{\nu,1}}$$

Theorem 7. *Theorem 14.19 in (**?**, p. 201). Assume we are dealing with models $\{P_{n,\theta} : \theta \geq 0\}$. Also assume that $\theta > 0$ and the total variation distance $\|P_{n,\theta} - P_{n,0}\| \to 0$ as $\theta \downarrow 0$. Let $T_{n,1}$ and $T_{n,2}$ be two statistics that satisfy*

$$\sqrt{n} \frac{T_{n,i} - \mu_i(\theta_n)}{\sigma_i(\theta_n)} \overset{\theta_n}{\rightsquigarrow} \mathcal{N}(0,1)$$

for $\theta_n \downarrow 0$, $\mu_i'(0) > 0$, $\sigma_i(0) > 0$, and σ_i continuous at 0. Then we have that the Pitman efficiency is equal to the ARE, i.e.

$$\lim_{\nu \to \infty} \frac{n_{\nu,2}}{n_{\nu,1}} = \left(\frac{\mu_1'(0)/\sigma_1(0)}{\mu_2'(0)/\sigma_2(0)} \right)^2$$

This is true for any $\theta_n \downarrow 0$ independent of α and γ.

Proof. First, we need to show that $n_{\nu_i} \to \infty$ for both i. This proof is given in the book (it arises from the assumption that the variation distance goes to zero). Since $n_{\nu,i} \to \infty$, we can use the assumption of asymptotic normality of $T_{n,i}$ to define tests of level α: we reject when

$$\sqrt{n_{\nu,i}}(T_{n_\nu,i} - \mu_i(0)) > \sigma_i(0) z_\alpha + o(1)$$

Then the power of this test is

$$\pi_{n_\nu,i}(\theta_\nu) = 1 - \Phi\left(z_\alpha + o(1) - \sqrt{n_{\nu,i}}\theta_\nu \mu_i'(0)/\sigma_i(0)(1+o(1))\right) + o(1)$$

where we have used the same derivative trick to get $\mu_i'(0)$ as before. This tends to level $\gamma < 1$ iff the arguments of Φ tends to z_γ. Therefore

$$\begin{aligned}
\lim_{\nu\to\infty} \frac{n_{\nu,2}}{n_{\nu,1}} &= \lim_{\nu\to\infty} \frac{n_{\nu,2}\theta_\nu^2}{n_{\nu,1}\theta_\nu^2} \\
&= \frac{(z_\alpha - z_\gamma)^2}{(\mu_2'(0)/\sigma_2(0))^2} \frac{(z_\alpha - z_\gamma)^2}{(\mu_1'(0)/\sigma_1(0))^2} \\
&= \left(\frac{\mu_1'(0)/\sigma_1(0)}{\mu_2'(0)/\sigma_2(0)}\right)^2 \\
&= \text{the ratio of the square of the slopes}
\end{aligned}$$

□

This has all been for 1-D cases. It gets harder in higher dimensional spaces. See the text for a flavor of higher dimensions.