University of Illinois — Department of Economics
Spring 2017 — Econ 574 — Roger Koenker

Lecture 12
Asymptotic Relative Efficiency of Tests: ARE on a G String

In this lecture I want delve a bit more deeply into the problem of comparing performance of various tests. Among other things I will try to provide an exposition of Chernoff's (1952) result employed in Lars Hansen's talks here several years ago. The lecture is based on van der Vaart (1998, Chapter 14).

Consider the problem of testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, using a test statistic T_n that rejects H_0 if T_n falls into the critical region K_n. The power function of the test is the function,

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

which describes how the probability of rejection depends upon the parameter θ, and the sample size, n. We say the test is of level α, or has size α, if

$$\sup_{\theta \in \Theta_0} \pi_n(\theta) \leq \alpha$$

For a given sample size, n, we say that a test with power function π_n is better than a test with power function $\tilde{\pi}_n$ if

$$\pi_n(\theta) \leq \tilde{\pi}_n(\theta) \qquad \theta \in \Theta_0$$

and

$$\pi_n(\theta) \geq \tilde{\pi}_n(\theta) \qquad \theta \in \Theta_1$$

and for some θ the inequalities hold strictly. These considerations simply require that the Type I and Type II errors be smaller for the $\pi_0(\theta)$ test than the $\tilde{\pi}_n(\theta)$ test. Unfortunately, this is very hard to verify, so we revert to asymptotic comparisons. The first strategy to suggest itself is to consider the limiting power function

$$\pi(\theta) = \lim_{n \to \infty} \pi_n(\theta)$$

Usually, however, this isn't really very informative as the next example indicates.

Example (Sign Test): Consider a random sample $X_1, \ldots, X_n$ from a distribution with unique median. The hypothesis $H_0 : \theta = 0$ can be tested against $H_1 : \theta > 0$ using the sign statistic

$$S_n = n^{-1} \sum_{i=1}^{n} I(X_i > 0)$$

If $F(x - \theta)$ is the *df* of the X's, then

$$ES_n = 1 - F(-\theta) \equiv \mu(\theta)$$

$$VS_n = (1 - F(-\theta))F(-\theta)/n = \sigma^2(\theta)$$

and using the normal approximation of the binomial

$$\sqrt{n}(S_n - \mu(\theta))/\sigma(\theta) \rightsquigarrow \mathcal{N}(0,1).$$

Under H_0, $\mu(0) = 1/2$ and $\sigma^2(0) = 1/4$ so we have

$$\sqrt{n}(S_n - 1/2) \rightsquigarrow \mathcal{N}(0, \frac{1}{4})$$

So a test that rejects H_0, if $T_n = \sqrt{n}(S_n - \frac{1}{2})$ exceeds the critical value $1/2 z_\alpha = 1/2\Phi^{-1}(1-\alpha)$ has the power function,

$$\begin{aligned} \pi_n(\theta) &= P_\theta(\sqrt{n}(S_n - \mu(\theta)) > \frac{1}{2}z_\alpha - (\mu(\theta) - \mu(0))) \\ &= 1 - \Phi\left(\frac{\frac{1}{2}z_\alpha - \sqrt{n}(F(0) - F(-\theta))}{\sigma(\theta)}\right) + o(1). \end{aligned}$$

But if F has a unique median, $F(0) - F(-\theta) > 0$ for every $\theta > 0$, so for any sequence $\alpha_n \to 0$ sufficiently slowly

$$\pi_n(\theta) \to \begin{cases} 0 & \text{if } \theta = 0 \\ 1 & \text{if } \theta > 0, \end{cases}$$

thus we find that the limiting power function of the sign test is perfect in the sense that the error probabilities of both types can be seen to both tend to zero simultaneously.

A test is said to be *consistent*, if it has power function $\pi_n(\theta) \to 1$ for all $\theta \in \Theta_1$. The sign test is consistent in this sense, but many other tests, like the familiar t-test of H_0 are also consistent so we need a more revealing criterion. Last week we introduced Pitman's idea of considering sequences of alternatives local to H_0.

Example: Let's see how this works in the sign test case. Now we have $H_1 : \theta = \theta_n$ and

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

Since $\sigma(0) = 1/2$, we see that the level of the test,

$$\pi_n(0) \to \alpha.$$

Now consider the power. If $\theta_n \to 0$ slowly, say like $1/\log n$ we have that

$$\sqrt{n}(F(0) - F(-\theta_n)) = \sqrt{n}f(0)(\log n)^{-1} \to \infty$$

so $\pi_n(\theta_n) \to 1$ for such sequences. But if $\theta_n = h/\sqrt{n}$, then we have

$$\pi_n(\theta_n) \to 1 - \Phi(z_\alpha - 2hf(0))$$

and the asymptotic power function tends to 1 as $h \to \infty$.

This can be formulated somewhat more generally as follows. Suppose that we have a statistic S_n such that for all sequences of local alternatives $\theta = h/\sqrt{n}$ we have

$$(*) \qquad \frac{\sqrt{n}(S_n - \mu(\theta_n))}{\sigma(\theta_n)} \rightsquigarrow \mathcal{N}(0,1)$$

where this convergence in distribution must be demonstrated under the whole sequence of models corresponding to $\theta_n = h/\sqrt{n}$. This entails that under H_0, $\sqrt{n}(S_n - \mu(0)) \rightsquigarrow \mathcal{N}(0, \sigma^2(0))$ so a test of H_0 would reject if $\sqrt{n}(S_n - \mu(0)) > \sigma(0)z_\alpha$ and we would have

$$\pi_0(\theta_n) = P_{\theta_n}(\sqrt{n}(S_n - \mu(\theta_n)) > \sigma(0)z_\alpha - \sqrt{n}(\mu(\theta_n) - \mu(0)))$$

so for $\theta_n = h/\sqrt{n}$ we have, for differentiable μ,

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

The quantity $\mu'(0)/\sigma(0)$ is the slope of the test since it measures how the probabilities change with h. If we have two tests that both can be expressed in this way, clearly we can compare them based solely on their slopes.

Example: The slope of the sign test we have already seen is $2f(0)$. If we now compare this with the usual t-test of H_0, for which

$$\sqrt{n}(\frac{\bar{X}_n}{S} - \mu(\theta_n)) \rightsquigarrow \mathcal{N}(0, \sigma^2(\theta_n))$$

where $\mu(\theta) = \theta/\sigma$, so $\mu(h/\sqrt{n}) = n^{-1/2}h/\sigma$, and $\sigma(\theta) = 1$, so the slope of the t-test is simply $1/\sigma$. So to compare the t-test and the sign test we need to compare $2f(0)$ with σ. This is done for several choices of f in the Table below.

Distribution	ARE (sign-test)/t-test
logistic	$\pi^2/12$
Normal	$2/\pi$
Laplace	2
Cauchy	∞
Uniform	1/3

An important interpretation of the ratios of slopes given in the previous table can be formulated in terms of the required sample sizes needed to achieve a specified level α, and power γ. Suppose for the moment that we could compute the precise finite sample power functions for two competing tests and let n_ν denote the sample size required to achieve,

$$\pi_{n_\nu}(0) \leq \alpha \qquad \text{and} \qquad \pi_{n_\nu}(\theta_\nu) \geq \gamma$$

Clearly we prefer tests for which n_ν is smallest. If it exists, we will call

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

the asymptotic relative efficiency, or Pitman efficiency, or just efficiency of the two tests. If $ARE = 2$, it means that twice as many observations are needed by test 2, than by test 1 to achieve a given level of performance.

Before stating the next result we must introduce the notion measuring the distance between absolutely continuous measures P, Q by their total variation distance

$$||P - Q|| = \int |p - q| d\mu.$$

Recall that the total variation of an absolutely continuous function f is given by

$$TV(f) = \int |f'| d\mu.$$

Theorem: Consider models $P_{n,\theta}$ such that $||P_{n,\theta} - P_{n,0}|| \to 0$ as $\theta \to 0$, for all n. Let $T_{n,1}$ and $T_{n,2}$ satisfy $(*)$ for sequences $\{\theta_n\}$ tending to 0 with functions $\mu_i, \sigma_i \quad i = 1, 2$, and $\mu_i'(0) > 0, \sigma(0) > 0$ then

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

Proof: We first establish that $n_{\nu,i} \to \infty$, as $\nu \to \infty$. Let α, β denote that Type 1 and 2 probabilities and write

$$\begin{aligned} \alpha + \beta &= \int_{K_n} dP_{n,0} + \int_{K_n^c} dP_{n,\theta_\nu} \\ &= 1 + \int_{K_n} (P_{n,0} - P_{n,\theta_\nu}) d\mu_n \end{aligned}$$

To find K_n, we minimize and set the critical region, $K_n = \{P_{n,0} < P_{n,\theta_\nu}\}$. Substituting this back in, yields,

$$\alpha + \beta = 1 - \frac{1}{2} ||P_{n,0} - P_{n,\theta_\nu}||.$$

But the *rhs* tends to 1 for any bounded sequence of $n = n_\nu$.

Since $n_{\nu,i} \to \infty$, we can employ the normal approximation,

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

So the power of the tests tends to $\gamma < 1$ iff the expression inside $\Phi(\cdot)$ tends to z_γ, but this implies that

$$\lim_{\nu \to \infty} \frac{n_{\nu,2}}{n_{\nu,1}} = \frac{n_{\nu,2}\theta_\nu^2}{n_{\nu,1}\theta_\nu^2} = \frac{(z_\alpha - z_\gamma)^2 / (\mu_2'(0)/\sigma_2(0))^2}{(z_\alpha - z_\gamma)^2 / (\mu_1'(0)/\sigma_1(0))^2}$$

And the result follows. □

As usual, the multivariate version of this result follows the same line of argument except now the limiting form of the test is χ^2 and the ARE ratio of limiting sample sizes is simply the ratio of the non-centrality parameters. This can be formalized by replacing $(*)$ by the condition

$$(**) \qquad n(S_n - \mu(\theta_n))' \Omega(\theta_n)^{-1} (S_n - \mu(\theta_n)) \rightsquigarrow \chi_P^2$$

So our test would reject if

$$T_n = n(S_n - \mu(0))'\Omega_0^{-1}(S_n - \mu(0)) > \chi^2_{P,\alpha}$$

so

$$\pi_n(\theta_n) = 1 - G_P(\chi^2_{P,\alpha} - n(\mu(\theta_n) - \mu(0))'\Omega_0^{-1}(\mu(\theta_n) - \mu(0)))$$

and for $\theta_n = h/\sqrt{n}$ we have, again for differentiable $\mu(\cdot)$,

$$\pi_n\left(\frac{h}{\sqrt{n}}\right) \to 1 - G_P(\chi^2_{P,\alpha} - h'\Omega_0^{-1}h)$$

where G_P is the *df* of the χ^2_P *r.v.*

Although Pitman efficiency is most commonly used in econometrics, there are other possibilities as well. We might consider a general form of relative efficiency as

$$\lim_{\nu\to\infty} \frac{n_2(\alpha_\nu, \gamma_\nu, \theta_\nu)}{n_1(\alpha_\nu, \gamma_\nu, \theta_\nu)}$$

Pitman fixes α_ν and γ_ν, and lets θ_ν drift toward $\theta_0 = 0$. Bahadur efficiency considers

$$\lim_{\nu\to\infty} \frac{n_2(\alpha_\nu, \gamma, \theta)}{n_1(\alpha_\nu, \gamma, \theta)} = BARE$$

So the alternative and power are fixed and $\alpha_\nu \to 0$. Typically, BARE depends upon γ and θ but not on how $\alpha_\nu \to 0$. The underlying theory of BARE is fundamentally different than Pitman ARE. While Pitman ARE is based on familiar CLT considerations BARE is based on large deviation results.

The usual situation is as follows. Suppose we are testing $H_0 : \theta = 0$, by rejecting for large values of T_n, and that for all t,

(i)
$$-\frac{2}{n}\log P_0(T_n \geq t) \to e(t)$$

and

(ii)
$$T_n \to \mu_\theta \quad \text{under } P_\theta.$$

The function $H(t) = P_0(T_n \geq t)$ is called the observed significance level of the test. If we evaluate $H(t)$ at the random $t = T_n$ we have a random variable that is uniformly distributed, under H_0. This is potentially confusing since it seems that $P_0(T_n \geq T_n) = 1$, but recall that if X has *df* $F(x) = P(X \leq x)$, then the *r.v.* $U = F(X)$ is $U[0,1]$. For a fixed alternative, say θ, $H(t)|_{t=T_n} \to 0$ at an exponential rate. In particular, under (i. - ii.) we have,

$$-\frac{2}{n}\log P_0(T_n \geq t)_{|t=T_n} \overset{P_\theta}{\to} e(\mu(\theta))$$

and the quantity $e(\mu(\theta))$ is called the Bahadur slope of the test, and the ratio of two such slopes is the $BARE$.

The primary tool needed for evaluating $BARE$'s is the large-deviation results (i.). For sample means this follows from the following result due to Cramér (1938) and Chernoff (1952).

Theorem: Let $Y_1 \ldots, Y_n$ be *iid* with cumulant generating function K. Then for every t,

$$n^{-1} \log P(\bar{Y}_n \geq t) \to \inf_{u \geq 0}(K(u) - tu)$$

Proof: We can restrict attention to the case $t = 0$, since the cgf of $Y_i - t$ is $K(u) - ut$. By Markov's inequality, for every $u \geq 0$,

$$\begin{aligned} P(\bar{Y}_n \geq 0) &= P(e^{un\bar{Y}_n} \geq 1) && (e^{un})^{\bar{Y}_n} \geq (e^{un})^0 \\ &\leq Ee^{un\bar{Y}_n} && \text{(Markov)} \\ &= e^{nK(u)} && * \end{aligned}$$

Thus,

$$n^{-1} \log P(\bar{Y}_n \geq t) \leq \inf_{u \geq 0}(K(u) - tu).$$

The lower bound is a bit more complicated. We first dispense with some special cases:

1. If $P(Y_i < 0) = 0$, then $K(u)$ is increasing on $\Re$ and $\inf K(u) = 0$, attained at $u = 0$, but $n^{-1} \log P(\bar{Y} \geq 0) = 0$ for every n so all is well.
2. If $P(Y_i > 0) = 0$, then $K(u)$ is decreasing on $\Re$ with $K(\infty) = \log P(Y_i = 0) = n^{-1} \log P(\bar{Y}_n \geq 0)$ so the claim is justified in this case also.
3. Now consider the case in which $K(u)$ is finite for every $u \in \Re$. Then we can differentiate under the integral

$$K(u) = \log \int e^{uy} dF(y)$$

to obtain

$$K'(u) = \frac{\int y e^{uy} dF}{\int e^{uy} dF}$$

so

$$K'(0) = EY_1$$

Since the Y_i's take both positive and negative values (points 1.) and 2.) above allow us to dispense with the contrary case) we can conclude that $K(u) \to \infty$ as $u \to \pm\infty$, and thus [inf $K(u)$] is attained at a point u_0 such that $K'(u_0) = 0$.

Case 1: If $u_0 < 0$, convexity of $K(u)$ implies that K is nondecreasing on $[u_0, \infty)$. This means that K attains its minimum over $u \geq 0$ at 0, where $K(0) = 0$. But $EY_1 = K'(0) > K'(u_0) = 0$ and therefore $P(\bar{Y} \geq 0) \to 1$ by SLLN and limit of left side is 0, so again the claim of the theorem is again vindicated.

Case 2: If $u_0 \geq 0$, let $Z_1, \ldots, Z_n$ be *iid* $r.v$'s with

$$dP_Z(z) = e^{-K(u_0)} e^{u_0 z} P_Y(z)$$

$^*Ee^{un\bar{Y}_n} = Ee^{u\Sigma Y_i} = (M(u))^n, M(u) = e^{\log M(u)} = e^{K(u)}, (M(u))^n = e^{nK(u)}$.

then Z_1 has cgf $K(u_0+u)-K(u_0)$ since

$$\begin{aligned}\log Ee^{uZ} &= \log \int e^{uz} e^{-K(u_0)+u_0 z} dP_Y(z) \\ &= K(u_0+u)-K(u_0)\end{aligned}$$

and as before $EZ_1 = K'(u_0) = 0$. Thus, for every $\varepsilon > 0$,

$$\begin{aligned}P(\bar{Y}_n \geq 0) &= EI(\bar{Z}_n \geq 0)e^{-u_0 n \bar{Z}_n} e^{nK(u_0)} \\ &\geq P(0 \leq \bar{Z}_n \leq \varepsilon)e^{-u_0 n\varepsilon} e^{nK(u_0)}\end{aligned}$$

where the $P(\cdot)$ term on the lhs is bounded away from 0, by the fact that $EZ_1 = 0$. From this we may conclude that

$$\lim n^{-1}[\log P(0 \leq \bar{Z}_n \leq \varepsilon) - u_0 n\varepsilon + nK(u_0)] \geq K(u_0) - u_0\varepsilon$$

and since this is true for all $\varepsilon > 0$, it is also true for $\varepsilon = 0$. The proof is completed by a truncation argument that removes the restriction to finite $K(u)$. See van der Vaart, p. 206, for details.

We illustrate this result by reconsidering the likelihood ratio statistic. Let $Y = \log(p_\theta/p_{\theta_0})$ and note that Y has cgf

$$\begin{aligned}K(u) = \log Ee^{uY} &= \log Ee^{u\log(p_\theta/p_{\theta_0})} p_{\theta_0} d\mu \\ &= \log \int p_\theta^u p_{\theta_0}^{1-u} d\mu\end{aligned}$$

For $0 \leq u \leq 1, K(u)$ is finite and formally differentiating we have

$$K'(u) = \frac{\int p_\theta^u p_{\theta_0}^{1-u} \log(p_\theta/p_{\theta_0}) d\mu}{\int p_\theta^u p_{\theta_0}^{1-u} d\mu}$$

so

$$K'(1) = E_\theta \log(p_\theta/p_{\theta_0}) = KLIC(p_\theta, p_{\theta_0}) \equiv \mu(\theta)$$

Thus,

$$\begin{aligned}n^{-1}\log P(\bar{Y}_n \geq \mu(\theta)) &\to \inf_{u\geq 0}(K(u) - u\mu(\theta)) \\ &= K(1) - \mu(\theta)\end{aligned}$$

so the Bahadur slope, $e(\mu(\theta))$, is

$$\begin{aligned}-\frac{2}{n}\log P(\bar{Y}_n \geq \mu(\theta)) &\to -2(K(1) - \mu(\theta)) \\ &= 2\mu(\theta).\end{aligned}$$

since $K(1) = \log \int p_\theta d\mu = 0$. □

This is sometimes called Stein's lemma, and is one of many significant results Stein never published. It shows that the Bahadur slope of the likelihood ratio test is $2 \times KLIC$.

Rescaling Rates

The Pitman drift formulation in which $H_1 : \theta_n = h/\sqrt{n}$ is typical in parametric problems, here we consider some more general situations.

Lemma: The power function $\pi_n(\theta)$ of any test satisfies,

$$\pi_n(\theta) - \pi_n(\theta_0) \leq \frac{1}{2}||P_{n,\theta} - P_{n,\theta_0}||$$

where $||P - Q|| = \int |p - q| d\mu$ is the total variation distance between P and Q

Proof Let $\phi_n(\cdot)$ denote a test, i.e., an indicator function that rejects if $\phi_n = 1$ and $= 0$ otherwise. Then

$$\pi_n(\theta) - \pi_n(\theta_0) = \int \phi_n(x)(p_{n,\theta}(x) - p_{n,\theta_0}(x))d\mu(x)$$

This expression is maximized by choosing

$$\phi_n(x) = I(p_{n,\theta}(x) > p_{n,\theta_0}(x))$$

but for any pair of densities p, q

$$\int_{q>p} (q - p)d\mu = \frac{1}{2} \int |p - q| d\mu$$

since $\int (p - q)d\mu = 0$. □

There is a nice connection between TV distance and Hellinger distance. Recall

$$H^2(P, Q) = \int (\sqrt{p} - \sqrt{q})^2 d\mu = 2 - 2 \int \sqrt{p}\sqrt{q} d\mu$$

The last integral is sometimes called the Hellinger affinity. For product measures this extends nicely as,

$$H^2(P^n, Q^n) = 2 - 2(1 - \frac{1}{2}H^2(P, Q))^n$$

This follows from the prior expression and the use of Fubini to write,

$$\begin{aligned} A(P^n, Q^n) &= \int p^{n/2} q^{n/2} d\mu \\ &= (\int p^{1/2} q^{1/2} d\mu)^n \end{aligned}$$