ECON 481
LECTURE 6: Local Asymptotics

Ivan A. Canay
Northwestern University

Setup

- **Generic testing problem**: observe data $X_i, i = 1, \dots, n$ i.i.d. with distribution
$$P \in \mathbf{P} = \{P_\theta : \theta \in \Theta\}$$
and test
$$H_0 : \theta \in \Theta_0 \text{ versus } H_1 : \theta \in \Theta_1 \ .$$

- **Test function**: a function $\phi_n = \phi_n(X_1, \dots, X_n)$ that returns the probability of rejecting the null hypothesis after observing $X_1, \dots, X_n$.

- **Example**: ϕ_n might be the indicator function of a certain test statistic $T_n = T_n(X_1, \dots, X_n)$ being greater than some critical value $c_n(1-\alpha)$.

Definition (Pointwise asymptotically of level α)

The test is said to be (pointwise) asymptotically of level α if,

$$\limsup_{n\to\infty} E_\theta\left[\phi_n\right] \leqslant \alpha, \quad \forall \theta \in \Theta_0 \ .$$

- **Includes**: Wald tests, quasi-likelihood ratio tests, and Lagrange multiplier tests.

Symmetric Location Model

Question: given two different tests of the same null hypothesis, $\phi_{1,n}$ and $\phi_{2,n}$, both being (pointwise) asymptotically of level α. How can one choose between these two tests?

Symmetric Location Model

- Let P_θ be the distribution with density $f(x-\theta)$ on the real line (w.r.t. Lebesgue measure). Suppose further that (1) f is symmetric about 0 and that (2) it's median, 0, is unique.
- f is symmetric about 0 $\Rightarrow f(x-\theta)$ is symmetric about θ.
- We also have that $E_\theta[X] = \theta$ and $\text{med}_\theta[X] = \theta$.
- Finally, suppose that (3) the variance of P_0 is positive and finite; that is, $\sigma_0^2 = \int x^2 f(x)dx \in (0,\infty)$.

Testing Problem: $\Theta_0 = \{0\}$ and $\Theta_1 = \{\theta \in \mathbf{R} : \theta > 0\}$; i.e., we wish to test the null hypothesis

$$H_0 : \theta = 0 \text{ versus } H_1 : \theta > 0 \,.$$

How could we test this null hypothesis?

SLM: t-test

$$\phi_{1,n} = I\left\{\frac{\sqrt{n}\bar{X}_n}{\hat{\sigma}_n} > z_{1-\alpha}\right\}$$

SLM: Sign-test

$$\phi_{2,n} = I\left\{ \frac{2}{\sqrt{n}} \sum_{1 \leqslant i \leqslant n} \left(I\{X_i > 0\} - \frac{1}{2} \right) > z_{1-\alpha} \right\}.$$

A Naive Approach to Power

- It is natural to base comparisons of two different tests on their power functions.
- **Power function**: the function
$$\pi_n(\theta) = E_\theta[\phi_n] \, ,$$
i.e., the probability of rejecting the null hypothesis as a function of the unknown parameter θ.
- In this problem it will be difficult to compare the finite-sample power functions of the two tests
- We may try to do so in an asymptotic sense.
- To this end, let's compute the power functions of each of the above two tests at a fixed $\theta > 0$.

A Naive Approach: t-test

Power function: $\pi_{1,n}(\theta) = E_\theta[\phi_{1,n}]$

A Naive Approach: sign-test

A Naive Approach: sign-test

$$S_n \equiv \frac{2}{\sqrt{n}} \sum_{1 \leqslant i \leqslant n} \left(I\{X_i > 0\} - \frac{1}{2} \right) = \frac{2}{\sqrt{n}} \sum_{1 \leqslant i \leqslant n} \Big(I\{X_i > 0\} - (1 - F(-\theta)) \Big) + 2\sqrt{n} (\frac{1}{2} - F(-\theta)) \,.$$

QUESTIONS?

Local Asymptotic Power

- There are an innumerable number of ways of embedding our situation with a sample of size n in a sequence of hypothetical situations with sample sizes larger than n.
- **Keep in mind**: we are really interested in the finite-sample behavior of the power function
- In our sample of size n we know that the power is not 1 uniformly for $\theta > 0$. May be very close to 1 for θ "far" from 0, but for θ "close" to 0 we would expect the finite-sample power function to be < 1.
- What we mean by "far" and "close" will change with our sample size n.

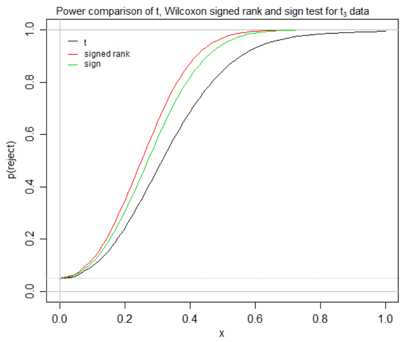

FIGURE: Exact Power Function

Local Asymptotic Power

- **Local Asymptotic Approximation**: considers the behavior of the power function evaluated at a sequence of alternatives θ_n, where θ_n tends to 0 (the null) at some rate. This provides a local (to the null) asymptotic approximation to the power function.

- If θ_n tends to 0 slowly enough, then the power function will still tend to 1 as n tends to infinity.

- If θ_n tends to 0 quickly enough, then for asymptotic purposes it's as if $\theta_n = 0$. For any such sequence, the power function tends to α as n tends to infinity in each of the above two examples.

- **Delicate rate** in between the two extremes above such that if θ_n tends to 0 at this rate, then the power will tend to a limit in $(\alpha, 1)$. This rate may be different in different problems, but in problems such as this one in which the distribution depends on θ in a "smooth" way it must be that

$$\theta_n = O\left(\frac{1}{\sqrt{n}}\right) .$$

- We will consider sequences $\theta_n = \frac{h}{\sqrt{n}}$, where $h \in \mathbf{R}$.

Local Power: t-test

Let $X_{i,n}, i = 1, \ldots, n$ be i.i.d. with distribution P_{θ_n} and let $Y_{i,n} = X_{i,n} - \theta_n \sim P_0$.

Local Power: sign-test

Start by studying $\frac{1}{n} \sum_{1 \leqslant i \leqslant n} I\{X_{i,n} > 0\}$ under P_{θ_n} .

Lindeberg-Feller Central Limit Theorem

Theorem

For each n, *let* $Z_{n,i}, i = 1, ..., n$ *be i.i.d. with distribution* P_n. *Suppose* $E_n[Z_{n,i}] = 0$ *and* $V_n[Z_{n,i}] = \sigma_n^2 < \infty$. *If for each* $\epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} E_n\left[Z_{n,i}^2 I\left\{|Z_{n,i}| > \epsilon\sqrt{n}\sigma_n\right\}\right] = 0$$

then

$$\frac{\sqrt{n}\bar{Z}_{n,n}}{\sigma_n} \xrightarrow{d} N(0,1)$$

under P_n.

We would like to use this result to assert that

$$S_n(\theta_n) = \frac{1}{\sqrt{n}} \sum_{1\leqslant i\leqslant n} (I\{X_{i,n} > 0\} - (1 - F(-\theta_n)))$$

converges in distribution under P_{θ_n} to a normal distribution.

Lindeberg-Feller CLT Application

CLT states that

$$\frac{\sqrt{n}\bar{Z}_{n,n}}{\sigma_n} \xrightarrow{d} N(0,1) \quad \text{if} \quad \lim_{n\to\infty} \frac{1}{\sigma_n^2} E_n[Z_{n,i}^2 I\{|Z_{n,i}| > \epsilon\sqrt{n}\sigma_n\}] = 0\,.$$

Let $Z_{n,i} = I\{X_{i,n} > 0\} - (1 - F(-\theta_n))$ with $\sigma_n^2 = F(-\theta_n)(1 - F(-\theta_n))$.

Local Power: sign-test

Recall: $S_n = 2S_n(\theta_n) + 2\sqrt{n}\left(\frac{1}{2} - F(-\theta)\right)$ and $2S_n(\theta_n) = 2\sigma_n \frac{S_n(\theta_n)}{\sigma_n} \xrightarrow{d} N(0,1)$

Local Asymptotic Power: summary

$$\pi_{1,n}(\theta_n) \to 1 - \Phi\left(z_{1-\alpha} - \frac{h}{\sigma_0}\right) \quad \text{versus} \quad \pi_{2,n}(\theta_n) \to 1 - \Phi\left(z_{1-\alpha} - 2hf(0)\right).$$

- If $2f(0) > \frac{1}{\sigma_0}$: the sign test will be preferred to the t-test in a local asymptotic power sense.
- **Normal case**: If f is the normal density, the t-test should be uniformly most powerful for testing the null hypothesis.
  $\Rightarrow$ If we plug in the standard normal density for f, we find that the above analysis bears this out.
- If we consider distributions with "fatter" tails (i.e., Laplace), the situation is reversed.
- **Moral of this story**: if the underlying distribution is symmetric, then, the t-test, while preferred for many distributions, is not as robust as the sign test to "fat" tails.
- **Asymptotic Relative Efficiency**: defined as the square of the ratio of $2f(0)$ to $1/\sigma_0$,

$$ARE_{2,1} = 4f(0)^2\sigma_0^2.$$

QUESTIONS?

Limits of Local Power Approximations

$$\pi_{1,n}(\theta_n) \to 1 - \Phi\left(z_{1-\alpha} - \frac{h}{\sigma_0}\right) \quad \text{versus} \quad \pi_{2,n}(\theta_n) \to 1 - \Phi\big(z_{1-\alpha} - 2hf(0)\big) .$$

- The local power function is monotonic and it has essentially the same shape as the power function in the normal location model.
- However: the accuracy of the approximation can be poor at non-local alternatives.
- **Non-monotonicity**: If the finite sample power curve is non-monotone, the asymptotic local power approximation will be poor at non-local alternatives.
- We will consider one example of this phenomenon presented by Nelson and Savin (1990). Another one, perhaps empirically more relevant, is the one in Savin and Wurtz (1999).

Non-Monotonicity of the t-Test

- Consider the following simple model in which

$$P \in \mathbf{P} = \{P_\theta : \theta \in \Theta\} \text{ and } P_\theta = N(\exp(\theta), 1) \ .$$

- Suppose that $\Theta_0 = \{0\}$ and that $\Theta_1 = \{\theta : \theta < 0\}$.

- You can think of this as a simple case of the following non-linear regression model with normal errors,

$$X_i = \exp(Z_i \theta) + U_i \ ,$$

  where $\theta \in \mathbf{R}$, Z_i is a scalar exogenous variable, and $U_i \sim N(0, 1)$ i.i.d.

- Simplicity of exposition: focus on the scalar case where $X_i, i = 1, \ldots, n$ is i.i.d. with distribution P as above and wishes to test the null hypothesis $H_0 : \theta = \theta_0$.

Non-Monotonicity of the t-Test

The log-likelihood function for the model and its derivatives are the following:

$$\ell(\theta) = c - \frac{1}{2}\sum_{i=1}^{n}(X_i - \exp(\theta))^2 \; ,$$

$$s(\theta) = \sum_{i=1}^{n}(X_i - \exp(\theta))\exp(\theta) \; ,$$

$$H(\theta) = \sum_{i=1}^{n}(X_i - 2\exp(\theta))\exp(\theta) \; .$$

The Fisher's information is then

$$I(\theta) = -E[H(\theta)] = (\exp(\theta))^2 \; ,$$

and the MLE of θ is just $\hat{\theta} = \log(\bar{X}_n)$.

t-STATISTIC

$$t(\hat{\theta}) = -\frac{\sqrt{n}(\hat{\theta} - \theta_0)}{I(\hat{\theta})^{-1/2}} = \sqrt{n}(\theta_0 - \hat{\theta})\left[(\exp(\hat{\theta}))^2\right]^{1/2} = \sqrt{n}(\theta_0 - \hat{\theta})\exp(\hat{\theta}) \; .$$

Non-Monotonicity of the t-Test

$$t(\hat{\theta}) = \sqrt{n}(\theta_0 - \hat{\theta}) \exp(\hat{\theta}) \ .$$

- t-statistic is a **non-monotonic function** of $\hat{\theta}$

$$\frac{\partial t(\hat{\theta})}{\partial \hat{\theta}} = \sqrt{n} \exp(\hat{\theta})[\theta_0 - \hat{\theta} - 1] \ ,$$

- Has a maximum value of

$$\sqrt{n} \exp(\theta_0 - 1) \text{ at } \hat{\theta} = \theta_0 - 1 \ .$$

- Decreases to zero on the left and $-\infty$ on the right.

- **Figure**: suggests that for any sample size the more negative the estimate of θ, the less likely the null hypothesis will be rejected in favor of the alternative that θ is negative!

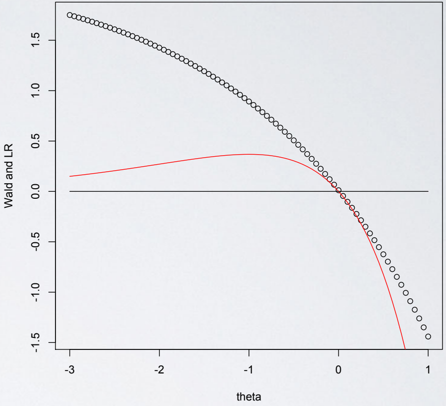

FIGURE: Wald test (line) versus signed $\sqrt{LR}$ (dots)

Power Function of the t-Test

- **Power function**:

$$\begin{aligned}\pi_n(\theta) &= P_\theta\left\{t(\hat{\theta}) > c_n(1-\alpha)\right\}\\ &= P_\theta\left\{\theta_L < \hat{\theta} < \theta_H\right\}\\ &= P_\theta\left\{\exp(\theta_L) < \bar{X}_n < \exp(\theta_H)\right\}\ ,\end{aligned}$$

  where θ_L and θ_H are the unique solutions to

$$t(\theta_L) = t(\theta_H) = c_n(1-\alpha)\ .$$

- non-monotonicity: two such values for any $c_n(1-\alpha)$ in the interval $(0, \sqrt{n}\exp(\theta_0 - 1))$.
- **Exact power**: $\bar{X}_n \sim N(\exp(\theta), n^{-1})$.
- Exact power approaches one as the true value of θ falls from 0 to about -0.85 and then declines for smaller values of θ.

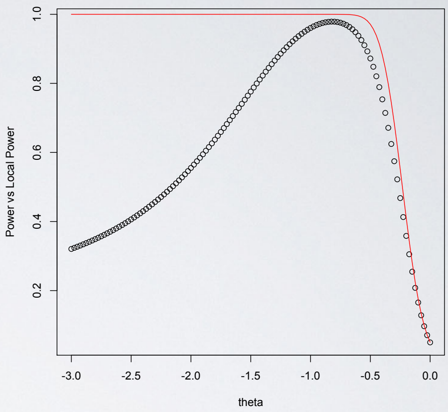

FIGURE: Exact Power (dots) versus local approximation (line)

Power in Binary Choice Models

- Most hypotheses tested in binary response models are composite: the null hypothesis restricts only a subset of the parameters. The remaining parameters are referred to as nuisance parameters.
- **Example**: one of the slope coefficients is zero.
- **Savin and Wurtz (99)**: show that for any fixed sample size, the power goes to zero along a particular sequence of alternatives that often occur in practice.
- The result applies to any non-randomized test with size less than one, and is derived for a finite sample.
- Therefore: the usual asymptotic results hold meaning that consistent tests can have non-monotonic power (in finite samples) for the sequence of alternatives of interest.

THE END!