Non-nested Hypotheses Testing

Methods Overview

Lerong Li

UTSPH

October 9, 2012

Review of LR

- LR: $\lambda(x) = \frac{sup\{f(y|\theta,\theta \in \Theta_0)\}}{sup\{f(y|\theta,\theta \in \Theta)\}}$
- Asymptotically, $-2log\lambda(x)$ is χ^2 with df = reduction in the size of parameter space after imposing restriction.
- If non-nested, the parameter spaces and likelihoods are unrelated.
- Hard to figure out the distritution under null model.

Non-nested Model

Definition

Two Models, say H_f and H_g, are said to be non-nested if it is not possible to derive one from the other either by means of parametric restriction or limiting process.

- Strictly Non-nested
- Partially Non-nested(overlapping)

Example

$$
\begin{aligned}
H_1: \quad & y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon_1 \\
H_2: \quad & y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_2 \\
H_3: \quad & y = \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon_3 \\
H_4: \quad & y = \beta_4 x_4 + \beta_5 x_5 + \varepsilon_4
\end{aligned}
$$

Example

Example

(Cox, 1961, Unconditional Models)

$H_f: \quad f(y|\theta) = \frac{1}{y(2\pi\theta_2)} exp[-\frac{(logy-\theta_1)^2}{2\theta_2}]$ log-normal

$H_g: \quad g(y|\gamma) = \frac{1}{\gamma} exp(-y/\gamma)$ exponential

Example

(Conditional Models)

$H_f: \quad y = X\alpha + \varepsilon_f, \qquad \varepsilon_f \sim N(0, \sigma^2 I)$

$H_g: \quad y = Z\beta + \varepsilon_g, \qquad \varepsilon_g \sim N(0, \omega^2 I)$

Formal Definition

- Introduce a statistical measure of the "closeness" betweeen two models.
- Kullback-Leibler information criteria (KLIC)

$$
\begin{aligned}
I_{fg}(\theta,\gamma) &= E_f[logf(y|\theta) - logg(y|\gamma)] \\
&= \int_{R_f} log\{\frac{f(y|\theta)}{g(y|\gamma)}\}f(y|\theta)dy
\end{aligned}
$$

  - $I_{fg}(\theta,\gamma)$ is KLIC measure of H_g w.r.t H_f
  - Not a distance measure, just mean information for discrimination in favor of $f(y|\theta)$ against $g(y|\gamma)$.
  - Invariant to transformation of θ and γ
  - $I_{fg}(\theta,\gamma) \geq 0$ with "=" hold iff $f = g$
  - Additive for iid sample.
- Assume the true model is H_h, θ_0 is true value under H_f, and γ_0 is true value under H_g
- Closeness measure of H_g to H_f is defined as $C_{fg}(\theta_0) = I_{fg}(\theta_0, \gamma_*(\theta_0))$ where $\gamma_*(\theta_0) = argmaxE_f(logg(\gamma))$
- Similarly, $C_{gf}(\gamma_0) = I_{gf}(\gamma_0, \theta_*(\gamma_0))$ where $\theta_*(\gamma_0) = argmaxE_g(logf(\theta))$

Formal Definition

Definitions

- (Nested) H_f is nested within H_g iff $C_{fg}(\theta_0) = 0 \ \forall \ \theta_0 \in \Theta$ and $C_{gf}(\gamma_0) \neq 0$ for some $\gamma_0 \in \Gamma$.
- (Strictly Non-nested) H_f and H_g are strictly non-nested iff $C_{fg}(\theta_0), C_{gf}(\gamma_0)$ are both non-zero $\forall \ \theta_0 \in \Theta, \gamma_0 \in \Gamma$
- (Partially non-nested)H_f and H_g are partially non-nested if $C_{fg}(\theta_0), C_{gf}(\gamma_0)$ are both non-zero for some $\theta_0 \in \Theta, \gamma_0 \in \Gamma$.

Example

Example

(Cox, 1961, revisit: log-normal vs exponential)

- $\gamma_*(\theta_0) = argmaxE_f(logg(\gamma))$ where $E_f(logg(\gamma)) = E_f[-log\gamma - y/\gamma] = -log\gamma - exp(\theta_{10} + 0.5\theta_{20})/\gamma$
- $\Longrightarrow \gamma_*(\theta_0) = exp(\theta_{10} + 0.5\theta_{20})$
- $\Longrightarrow log\frac{f(y|\theta_0)}{g(y|\gamma_\star(\theta_0))} =$ $-0.5log(2\pi\theta_{20}) - (logy - \theta_{10})^2/2\theta_{20} - logy + \theta_{10} + 0.5\theta_{20} + yexp(\theta_{10} + 0.5\theta_{20})$
- $\Longrightarrow C_{fg} = E_f\{log\frac{f(y|\theta_0)}{g(y|\gamma_\star(\theta_0))}\} = -0.5log(2\pi\theta_{20}) + 0.5 + 0.5\theta_{20} \neq 0$
- Similarly to get $C_{gf} \neq 0$
- H_f and H_g are strictly non-nested

Cox Statistic

Let $y_i \quad i = 1 : n$ be iid random sample with p.d.f. $f(y|\theta)$ under H_f and $g(y|\gamma)$ under H_g, where $f(y|\theta)$ and $g(y|\gamma)$ are non-nested model.

- Test: $H_f : \quad y \sim f(y|\theta) \qquad vs \qquad H_g \quad y \sim g(y|\gamma)$
- Cox statistic (1961,1962):

$$T_f = log \frac{L_f(\hat{\theta}|y)}{L_g(\hat{\gamma}|y)} - E_f[log \frac{L_f(\hat{\theta}|y)}{L_g(\hat{\gamma}|y)}]$$

- Standardized cox statistic:

$$N_f = \frac{T_f}{\sqrt{var(T_f)}} \sim N(0,1)$$

Notes

- N_f is for the test of H_f against H_g.
- N_g is calculated similarly as $\frac{T_g}{\sqrt{var(T_g)}} \sim N(0,1)$
- Decision rule:

		N_f Reject	N_f Accept
N_g	Reject	$\lvert N_f\rvert > C_\alpha; \lvert N_g\rvert > C_\alpha$	$\lvert N_f\rvert \leq C_\alpha; \lvert N_g\rvert > C_\alpha$
N_g	Accept	$\lvert N_f\rvert > C_\alpha; \lvert N_g\rvert \leq C_\alpha$	$\lvert N_f\rvert \leq C_\alpha; \lvert N_g\rvert \leq C_\alpha$

- Difficulties of computation in expectation and variance.
  - Using KLIC to approximate expectation
  - se: $n^{-1}d'[I_n - R(\hat{\theta})[R'(\hat{\theta})R(\hat{\theta})]^{-1}R'(\hat{\theta})]d$
- Weakness or strength?

Some related methods

- Comprehensive Approach
  - Consider a comprehensive model including H_g and H_f as special cases.
  - For example, $[f(y|\theta)]^{\lambda}[g(y|\gamma)]^{1-\lambda}$, test about $\lambda = 0 \quad vs \quad \lambda = 1$
- Encompassing Approach
  - Consider whether model H_f can explain one or more features of the rival model H_g.
  - Assume unknown true model H_h, and define pseudo-true values

$$\theta_{h\star} = argmaxE_h[logf(y|\theta)], \qquad \gamma_{h\star} = argmaxE_h[logg(y|\gamma)]$$

  - Define H_f encompass H_g by $\gamma_{h\star} = \gamma_\star(\theta_{h\star})$
  - $\sqrt{n}(\hat{\gamma} - \gamma_\star(\hat{\theta})) \sim N(0, V(f,g))$
- References(Pesaran,1999)

Motivation

- What if both models are rejected or fail to be rejected?
- Select the best one.
- Test the hypothesis that models under consideration are equally close to the true model H_h.
- Closeness measure between hypothesis models and true model

$$C_{hf}(\theta_{h\star}) = E_h\{logh(y|.) - logf(\theta_{h\star})\}$$

$$C_{hg}(\gamma_{h\star}) = E_h\{logh(y|.) - logg(\gamma_{h\star})\}$$

Vuong's test(1989)

Null hypothesis underlying Vuong's approach is given by

$$H_V : \quad C_{hf}(\theta_{h\star}) = C_{hg}(\gamma_{h\star}) \iff$$

$$H_v : \quad E_h\{logf(\theta_{h\star})\} = E_h\{logg(\gamma_{h\star})\} \iff$$

$$H_v : \quad E_h\{log\frac{f(\theta_{h\star})}{g(\gamma_{h\star})}\} = 0$$

Alternative hypotheses are

$$H_f : \quad E_h\{log\frac{f(\theta_{h\star})}{g(\gamma_{h\star})}\} > 0 \qquad H_f \text{ is better}$$

$$H_g : \quad E_h\{log\frac{f(\theta_{h\star})}{g(\gamma_{h\star})}\} < 0 \qquad H_g \text{ is better}$$

Notes of Vuong's test

- Distribution of $E_h\{log\frac{f(\theta_{h\star})}{g(\gamma_{h\star})}\}$is unknown and depends on unknown true model H_h.
- Can be consistently estimated by $\frac{1}{n}log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}$
- Under null,by CLT: $\sqrt{n}(\frac{1}{n}log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}-0)\sim N(0,\hat{\omega}^2)$
- $\hat{\omega}^2 = v\hat{a}r(log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}) = \frac{1}{n}\Sigma[log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}]^2 - [\frac{1}{n}\Sigma log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}]^2$
- Adjustment of $log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})}$ if the number of coefficients in two models are different.
- $LR_{adj} = log\frac{f(y|\hat{\theta})}{g(y|\hat{\gamma})} - [\frac{p}{2}log(n) - \frac{q}{2}log(n)]$, where p and q are the number of estimated coefficients in models f and g respctively.

Review

Parameters are random variables

- Parameter θ drawn from prior $p(\theta)$
- Data D garthered conditionally on some fixed θ
- Update the prior $p(\theta)$ with garthered data D

$$P(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

- Drop marginal $p(D)$ since constant

$$p(\theta|D) \propto p(D|\theta) \times p(\theta)$$

Model Selection

- Instead of parameters, we consider models M.

$$P(M|D) = \frac{p(D|M)p(M)}{p(D)} \tag{1}$$

where $p(D|M) = \int_\theta p(D|\theta, M)p(\theta|M)d\theta$, marginalized over parameters space.

- Now Consider two models, M_f and M_g.From equation (1), we can have posterior odds in favor of M_f against M_g

$$\underbrace{\frac{p(M_f|D)}{p(M_g|D)}}_{Posterior\ odds} = \underbrace{\frac{p(D|M_f)}{p(D|M_g)}}_{Bayes\ factor\ K} \times \underbrace{\frac{p(M_f)}{p(M_g)}}_{Prior\ odds}$$

- Often assume $p(M_f) = p(M_g) = 1/2$, prior odds=1 , so Posterior odds = Bayes Factor

Decision Rule(Jeffreys,1961)

$$
\begin{aligned}
K>1 &\Longrightarrow \text{Null hypothesis supported} \\
K<1 &\Longrightarrow \text{Not worth more than bare mention} \\
K<1/\sqrt{10} &\Longrightarrow \text{Evidence against } M_f \text{ substantial} \\
K<1/10 &\Longrightarrow \text{Strong} \\
K<1/10^{3/2} &\Longrightarrow \text{very strong} \\
K<1/100 &\Longrightarrow \text{Evidence against } M_f \text{ decisive}
\end{aligned}
$$

Notes

- Hard to evaluate

$$p(D|M) = \int_\theta p(D|\theta, M)p(\theta|M)d\theta$$

- Approximation
    - BIC (Schwarz,1976)
    - MCMC (Smith,1997)
    - Laplace approximation (Jeffrey,1961)

Other Methods

- Distribution free test(Clarke,2003)
- Boostrap statistic

Reference

- D.R.COX, (1961)"Tests of separate families of hypotheses"
- D.R.COX, (1962) "Further results on tests of separate families of hypothese"
- QUANG H. VUONG (1989), "Likelihood ratio tests for model selection and non-nested hypotheses"
- M.H.Pesaran& M.Weeks (1999) "Non-nested hypothesis testing: An overview"
- Kevin A. Clarke (2001)"Testing Nonnested models of international relations: Reevaluating Realism"
- Kevin A. Clarke (2003) "A simple distribution-free test for non-nested hypotheses"

THANKS