Econometrics I, Testing

Department of Economics
Stanford University

September, 2008

Part I

- Null hypothesis H_0. Alternative hypothesis H_A.
- Sample $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.
- Rejection region: $R \in \mathbb{R}^n$ such that H_0 is rejected if $\mathbf{x} \in R$.
- Simple hypothesis; composite hypothesis.
    - Simple H_0 vs simple H_A;
    - Simple H_0 vs Composite H_A;
    - Composite H_0 vs Composite H_A.

Principle of testing (of finding R):

- Find a scalar statistic $t(\mathbf{x})$ such that under the null H_0:

$$t(\mathbf{x}) \xrightarrow{p} 0$$

while under the alternative H_A:

$$t(\mathbf{x}) \xrightarrow{p} c > 0.$$

- Derive (asymptotic or exact) distribution of $t(\mathbf{x})$ under H_0:

$$\text{Under } H_0\text{:} \quad P(a_n t(\mathbf{x}) \leq x) \overset{A}{\sim} F(x)$$

where $a_n \to \infty$ when $n \to \infty$

- Reject H_0 if $a_n t(\mathbf{x})$ is larger than αth critical value of $F(x)$.
- Under H_A: $a_n t(\mathbf{x}) \to \infty$.

Trinity of tests

- Likelihood ratio test.
- Need to estimate model under both H_0 and H_A.
- Wald test.
- No need to estimate under H_0.
- Score function test.
- Only need to estimate under H_0.

Properties of testings

- Type I error: error of rejecting H_0 when it is true.
- Size: probability of type I error (under H_0 by definition).

$$P\left(a_n t\left(\mathbf{x}\right) \geq F^{-1}\left(1-\alpha\right)\right) \approx \alpha \quad \text{when } H_0 \text{ is true.}$$

- Type II error: error of accepting H_0 when H_A is true.
- Power: probability of rejecting H_0 when H_A is true.
- Therefore, power = 1 - P(type II error) = 1 - β.
- Power of asymptotic test is usually 1:

$$P\left(a_n t\left(\mathbf{x}\right) \geq F^{-1}\left(1-\alpha\right)\right) \rightarrow 1 \quad \text{when } H_A \text{ is true.}$$

- Consistent test. Local asymptotic power.

Simple null versus simple alternative

- Definition 9.2.2: Let (α_1, β_1) and (α_2, β_2) be the characteristics of two tests. The first test is better (more powerful) than the second test if $\alpha_1 \leq \alpha_2$ and $\beta_1 \leq \beta_2$ with a strict inequality holding for at least one of the $\leq$.
- Definition 9.2.4: R is the most powerful test of size α if $\alpha(R) = \alpha$ and for any test R_1 of size α, $\beta(R) \leq \beta(R_1)$. (It may not be unique.)
- Definition 9.5.2: R is the most powerful test of level α if $\alpha(R) \leq \alpha$ and for any test R_1 of level α (that is, such that $\alpha(R_1) \leq \alpha$, $\beta(R) \leq \beta(R_1)$.
- "level α" is needed with discrete sample when $\alpha(R) \neq \alpha$ for all R. Not needed with randomized test.
- Randomized test: toss a coin before deciding which of R_1 and R_2 to use. Can achieve any desired α.

Likelihood Ratio Test

- simple H_0 versus simple H_1: $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, reject if

$$\frac{L(x|\theta_1)}{L(x|\theta_0)} > c \quad \text{or} \quad R = \left\{x : \frac{L(x|H_1)}{L(x|H_0)} > c\right\}$$

- Simple $H_0 : \theta = \theta_0$, composite H_1, for example, $H_1 : \theta > \theta_0$,

$$LR = \frac{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(x|\theta)}{L(x|\theta_0)} = \frac{L\left(X|\hat{\theta}_{MLE}\right)}{L(X|\theta_0)} \geq 1.$$

  $H_0 \cup H_1 = \{\theta : \theta \geq \theta_0\}$ parameter space, over which we calculate MLE.

- Composite H_0, composite H_1. Reject if

$$LR = \frac{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(x|\theta)}{\sup_{\theta \in \Theta_0} L(x|\theta) > c}$$

  use $H_0 \cup H_1$ for convenience because this is just unconstrained MLE.

- In simple H_0 versus simple H_1 tests, the LR test is the most powerful test (of a given size),
- Bayes testing: loss matrix,

	State of Nature	
Decision	H_0	H_1
H_0	0	γ_2
H_1	γ_1	0

- Bayesian expected loss

$$\begin{aligned}\phi(R) = & \underbrace{\gamma_1 \pi(H_0)}_{\eta_0} P(\mathbf{x} \in R | H_0) + \underbrace{\gamma_2 \pi(H_1)}_{\eta_1} P(\mathbf{x} \in R^c | H_1) \\ \equiv & \eta_0 \alpha(R) + \eta_1 \beta(R) \\ = & \eta_0 \int_{\mathbf{x} \in R} f(\mathbf{x}|H_0)\, d\mathbf{x} + \eta_1 \int_{\mathbf{x} \in R^c} f(\mathbf{x}|H_1)\, d\mathbf{x}.\end{aligned}$$

- Optimal R that minimizes $\phi(R)$:

$$R_0 = \{\mathbf{x} : \eta_0 f(\mathbf{x}|H_0) \leq \eta_1 f(\mathbf{x}|H_1)\} = \{\mathbf{x} : \frac{L(\mathbf{x}|H_1)}{L(\mathbf{x}|H_0)} > \frac{\eta_0}{\eta_1}\}$$
$$= \{\mathbf{x} : \log L(\mathbf{x}|H_1) - \log L(\mathbf{x}|H_0) > \log\left(\frac{\eta_0}{\eta_1}\right)\}.$$

- Likelihood ratio test minimizes Bayes risk.
- is also most powerful.
- $\phi(R)$ is a linear combination of size and type II error.
- No R_1 s.t. $\alpha(R_1) = \alpha(R_0)$ and $\beta(R_1) < \beta(R_0)$.
- Otherwise $\phi(R_1) < \phi(R_0)$.
- Frequentist: pick η_0/η_1 for the desired size.

- Example. $H_0 : \mu = \mu_0$, $H_A : \mu = \mu_1$. Assume $\mu_1 > \mu_0$.
- $\{X_t, t = 1, \ldots\}$ i.i.d. $N\left(\mu, \sigma^2\right)$. σ^2 known.
- Log likelihood ratio:

$$\begin{aligned} & -\frac{1}{2\sigma^2} \sum \left(x_t - \mu_1\right)^2 + \frac{1}{2\sigma^2} \sum \left(x_t - \mu_0\right)^2 \\ = & \frac{n\left(\mu_1 - \mu_0\right)}{\sigma^2} \bar{x} + c \end{aligned}$$

- Best rejection region: $\bar{x} > d$ such that

$$P\left(\bar{x} > d | H_0\right) = \alpha.$$

- Note that d does not depend on the value of μ_1, as long as

$$\mu_1 > \mu_0$$

- Under the H_0:

$$
\begin{aligned}
P\left(\bar{x} > d | H_0\right) =& P\left(\frac{\sqrt{n}\left(\bar{x} - \mu_0\right)}{\sigma} > \frac{\sqrt{n}\left(d - \mu_0\right)}{\sigma}\right) \\
=& P\left(Z > \frac{\sqrt{n}\left(d - \mu_0\right)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}\left(d - \mu_0\right)}{\sigma}\right) = \alpha
\end{aligned}
$$

Solve for d,

$$
d = \Phi^{-1}\left(1 - \alpha\right)\frac{\sigma}{\sqrt{n}} + \mu_0
$$

- Simple Null vs Composite Alternative.
- $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta_1$.
- Power function for a test R

$$Q(\theta) = P(\mathbf{x} \in R|\theta).$$

- $Q(\theta_0) = \alpha$. $Q(\theta_1) = 1 - \beta$ for $\theta_1 \in \Theta_1$.
- Definition 9.4.2. R_1 is uniformly better than R_2 if $Q_1(\theta_0) = Q_2(\theta_0)$, $Q_1(\theta) \geq Q_2(\theta)$ $\forall \theta \in \Theta_1$, and $Q_1(\theta_1) > Q_2(\theta_1)$ for at least one $\theta_1 \in \Theta_1$.
- Definition 9.4.3. A test R is uniformly most powerful (UMP) if it is uniformly weakly better than any other test with the same size (the same $Q(\theta_0)$).

- UMP tests often do not exist.
- Likelihood ratio test usually is UMP if UMP exists.
- Likelihood ratio test often used even without UMP.
- LR test: reject if

$$\log L(\theta_0) - \sup_{\theta \in \theta_0 \cup \Theta_1} \log L(\theta) < c.$$

- The second part is just the maximum likelihood estimator:

$$\log L\left(\hat{\theta}\right) = \sup_{\theta \in \theta_0 \cup \Theta_1} \log L(\theta).$$

- An admissible test is a test R_0 such that there is no other test R with the same size, where $PR(R|\theta) \geq PR(R_0|\theta)$ for all θ, and $PR(R|\theta) > PR(R_0|\theta)$ for some θ.

- Example. $H_0 : \mu = \mu_0$, $H_A : \mu > \mu_0$.
- $\{X_t \sim N\left(\mu, \sigma^2\right)\}$ with σ^2 known.
- log likelihood ratio statistics:

$$
\begin{aligned}
LR &= -\frac{1}{2\sigma^2}\sum\left(x_t - \mu_0\right)^2 - \sup_{\mu \geq \mu_0} -\frac{1}{2\sigma^2}\sum\left(x_t - \mu\right)^2 \\
&= -\frac{1}{2\sigma^2} n\left(\bar{x} - \mu_0\right)^2 + \inf_{\mu \geq \mu_0} \frac{1}{2\sigma^2} n\left(\bar{x} - \mu\right)^2 .
\end{aligned}
$$

- If $\bar{x} \leq \mu_0$, then $\inf_{\mu \geq \mu_0} \frac{1}{2\sigma^2} n\left(\bar{x} - \mu\right)^2 = \frac{1}{2\sigma^2} n\left(\bar{x} - \mu_0\right)^2$. $LR = 0$.
- Do not reject.
- If $\bar{x} > \mu_0$, $\inf = 0$, $LR = -\frac{1}{2\sigma^2} n\left(\bar{x} - \mu_0\right)^2$, reject if $\bar{x} > d$.
- This test is UMP.

- Same test for simple $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$ for any $\mu_1 > \mu_0$, same d as before. Reject if

$$\bar{x} > d = \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}} + \mu_0.$$

- Power function, for $\mu > \mu_0$,

$$\begin{aligned}
Power(\mu) =& P(\bar{x} > d|\mu) = P\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma} \geq \Phi^{-1}(1-\alpha)|\mu\right) \\
=& P\left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \geq \Phi^{-1}(1-\alpha) + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}|\mu\right) \\
=& P\left(Z \geq \Phi^{-1}(1-\alpha) + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right) \\
=& 1 - \Phi\left(\Phi^{-1}(1-\alpha) + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right)
\end{aligned}$$

$Power(\mu) \to 1$ as $n \to \infty$, or $\mu \to \infty$.

- Example: $H_0 : \mu = \mu_0$, $H_A : \mu \neq \mu_0$.
- $\{X_t \sim N\left(\mu, \sigma^2\right)\}$ with σ^2 known.
- log likelihood ratio statistics:

$$\begin{aligned} LR &= -\frac{1}{2\sigma^2}\sum\left(x_t - \mu_0\right)^2 - \sup_{\mu\neq\mu_0} -\frac{1}{2\sigma^2}\sum\left(x_t - \mu\right)^2 \\ &= -\frac{1}{2\sigma^2} n\left(\bar{x} - \mu_0\right)^2 + \inf_{\mu\neq\mu_0} \frac{1}{2\sigma^2} n\left(\bar{x} - \mu\right)^2 \\ &= -\frac{1}{2\sigma^2} n\left(\bar{x} - \mu_0\right)^2 . \end{aligned}$$

- Reject if $|\bar{x} - \mu_0| > d$ where

$$P\left(|\bar{x} - \mu_0| > d | H_0\right) = \alpha.$$

- Not a UMP test. Not Neyman-Pearson against, e.g. any $\mu > \mu_0$. But, UMP among tests that assign equal power to μ equidistant from μ_0.

- To determine d,

$$P\left(\left|\frac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{\sigma}\right| \geq \frac{\sqrt{n}d}{\sigma}\right) = \alpha$$

$$d = \Phi^{-1}\left(1-\alpha/2\right)\frac{\sigma}{\sqrt{n}}.$$

- Power function, as $n \to \infty$ or $|\mu| \to \infty$.

$$\begin{aligned}
&P\left(|\bar{x}-\mu_0| > d|\mu\right) = P\left(\bar{x} > \mu_0 + d|\mu\right) + P\left(\bar{x} < \mu_0 - d|\mu\right) \\
&= P\left(\bar{x}-\mu > \mu_0 - \mu + d|\mu\right) + P\left(\bar{x}-\mu < \mu_0 - \mu + d|\mu\right) \\
&= P\left(\frac{\sqrt{n}\left(\bar{x}-\mu\right)}{\sigma} > \frac{\sqrt{n}\left(\mu_0-\mu+d\right)}{\sigma}|\mu\right) \\
&\quad + P\left(\frac{\sqrt{n}\left(\bar{x}-\mu\right)}{\sigma} < \frac{\sqrt{n}\left(\mu_0-\mu-d\right)}{\sigma}|\mu\right) \\
&= 1 - \Phi\left(\frac{\sqrt{n}\left(\mu_0-\mu\right)}{\sigma} + \Phi^{-1}\left(1-\alpha/2\right)\right) \\
&\quad + \Phi\left(\frac{\sqrt{n}\left(\mu_0-\mu\right)}{\sigma} - \Phi^{-1}\left(1-\alpha/2\right)\right) \to 1
\end{aligned}$$

- Composite H_0 vs Composite H_1: UMP test may exist.
- Example: $X \sim N(\mu, 1)$, $H_0 : \mu \leq 0$, $H_1 : \mu > 0$.
- Recall that $\alpha(R) = \sup_{\mu \in H_0} P(X \in R|\mu)$.
- The convention is not to use the "size function", but instead use the "least favorable null distribution".
- What if you reject if $X > \Phi^{-1}(1-\alpha)$. Then

$$
\begin{aligned}
\alpha(R) &= \sup_{\mu \leq o} P(X > Z_{1-\alpha}|\mu) \\
&= \sup_{\mu \leq 0} P(X - \mu > Z_{1-\alpha} - \mu) \\
&= \sup_{\mu \leq 0} (1 - \Phi(Z_{1-\alpha} - \mu)) = 1 - \Phi(Z_{1-\alpha}) = \alpha
\end{aligned}
$$

- By definition, any test of size α for composite H_0 against composite H_1 will also have size at most α at the least favorable null $\mu = 0$, and therefore the above test is UMP. Same as the one sided test of $H_0 : \mu = 0$ against $H_1 : \mu > 0$.

- Wald tests: $H_0 : \theta = \theta_0$. $H_A : \theta \neq \theta_0$.
- Typically $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \Sigma)$.
- Intend to reject if $\hat{\theta} - \theta_0$ is sufficient large.
- What metric to use to measure distance $|\hat{\theta} - \theta_0|$?
- Use quadratic norm:

$$a_n t(\mathbf{x}) = n\left(\hat{\theta} - \theta_0\right)' \hat{\Sigma}^{-1}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \chi^2_{\dim(\theta)}.$$

- If $\hat{\theta}$ is MLE: $\Sigma = H^{-1}SH^{-1}$, and if the model is also correctly specified, $\Sigma = -H^{-1} = S^{-1}$.
- Why use this particular weighting matrix? So that limit has no nuisance parameters.

Theorem: Suppose X is a J-vector distributed as $N(\mu, \Sigma)$, then $(X-\mu)' \Sigma^{-1} (X-\mu) \sim \chi^2_J$.
Proof: Σ is symmetric and positive definite. So use an eigen value-function decomposition:

$$\Sigma = H \Lambda H'$$

Λ is a diagonal matrix with positive characteris roots of Σ on the main diagonal. H is orthonormal: $HH' = I \Rightarrow H' = H^{-1}$. Let $\Sigma^{-1/2} = H \Lambda^{-1/2} H'$. Then

$$\begin{aligned} \Sigma^{-1/2}\Sigma^{-1/2} =& H\Lambda^{-1/2}H'H\Lambda^{-1/2}H' = H\Lambda^{-1/2}\Lambda^{-1/2}H' = H\Lambda^{-1}H', \\ \Sigma^{-1} =& \left(H\Lambda H'\right)^{-1} = H'^{-1}\Lambda^{-1}H^{-1} = H\Lambda^{-1}H' = \Sigma^{-1/2}\Sigma^{-1/2}. \\ \Sigma^{-1/2}\Sigma\Sigma^{-1/2} =& H\Lambda^{-1/2}H'H\Lambda H'H\Lambda^{-1/2}H' = HH' = I. \end{aligned}$$

Therefore

$$\Sigma^{-1/2}(X-\mu) \sim N\left(0, \Sigma^{-1/2}\Sigma\Sigma^{-1/2}\right) = N(0, I)$$

$$\left(\Sigma^{-1/2}(X-\mu)\right)'\left(\Sigma^{-1/2}(X-\mu)\right) \sim \chi^2_J.$$

- Wald test for linear combinations:
- $H_0 : A\theta = A\theta_0 = b$, $H_A : A\theta \neq A\theta_0 = b$.
- If $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \Sigma)$, then under the H_0:

$$\sqrt{n}\left(A\hat{\theta} - b\right) = \sqrt{n}A\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, A\Sigma A')$$

- Quadratic norm based test statistic

$$\begin{aligned} & n\left(A\hat{\theta} - b\right)'\left(A\hat{\Sigma}A'\right)^{-1}\left(A\hat{\theta} - b\right) \\ =& n\left(\hat{\theta} - \theta_0\right)' A'\left(A\hat{\Sigma}A'\right)^{-1} A\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \chi^2_{\text{rows of } A} \end{aligned}$$

- Not UMP.

- Wald test for nonlinear functions

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0,\Sigma\right)$$
$$H_0 : g\left(\theta_0\right)=0 \quad H_1 : g\left(\theta_0\right)\neq 0.$$

where $g\left(\theta\right)=\left(g_1\left(\theta\right),\ldots,g_J\left(\theta\right)\right)'$ is $J\times 1$

- By the Delta method,

$$\sqrt{n}\left(\hat{g}\left(\hat{\theta}\right)-g\left(\theta_0\right)\right) \xrightarrow{d} N\left(0, R\left(\theta_0\right)\Sigma R\left(\theta_0\right)'\right)$$

where the $J\times d_\theta$ matrix,

$$R\left(\theta_0\right)=\frac{\partial g\left(\theta_0\right)}{\partial\theta'}=\left[\begin{array}{ccc}\frac{\partial g_1(\theta_0)}{\partial\theta_1} & \ldots & \frac{\partial g_1(\theta_0)}{\partial\theta_{d_\theta}}\\ \ldots & \ldots & \ldots\\ \frac{\partial g_J(\theta_0)}{\partial\theta_1} & \ldots & \frac{\partial g_J(\theta_0)}{\partial\theta_{d_\theta}}\end{array}\right]$$

- Define $A = R(\theta_0)\Sigma R(\theta_0)'$. Under $H_0 : g(\theta_0) = 0$,

$$\sqrt{n} g\left(\hat{\theta}\right) \xrightarrow{d} N(0, A)$$

- Let $A = H\Omega H'$. Define $A^{-1/2} H\Lambda^{-1/2} H'$.

$$\sqrt{n} A^{-1/2} g\left(\hat{\theta}\right) \xrightarrow{d} N\left(0, A^{-1/2} A A^{-1/2}\right) = N(0, I)$$

by the continuous mapping theorem

$$\left(\sqrt{n} A^{-1/2} g\left(\hat{\theta}\right)\right)' \left(\sqrt{n} A^{-1/2} g\left(\hat{\theta}\right)\right) \xrightarrow{d} \chi^2_J$$

Therefore,

$$W = n g\left(\hat{\theta}\right)' A^{-1} g\left(\hat{\theta}\right) \xrightarrow{d} \chi^2_J$$

Reject if $W > \chi^2_{J,\alpha}$.

- Also need $\hat{A} \xrightarrow{p} A$: $\hat{A} = R\left(\hat{\theta}\right) \Sigma R\left(\hat{\theta}\right)$. Define

$$\hat{W} = ng\left(\hat{\theta}\right)' \hat{A}^{-1} g\left(\hat{\theta}\right) \xrightarrow{d} \chi^2_J$$

- Use a combination of Slutsky and CMT to show

$$\hat{W} \xrightarrow{d} \chi^2_J$$

- Asymptotic LR test: $H_0 : \theta = \theta_0$, $H_A : \theta \neq \theta_0$.

$$LR = \log L(\theta_0) - \log L\left(\hat{\theta}_{MLE}\right).$$

- One can prove that under H_0,

$$2LR \xrightarrow{d} -\chi^2_{\dim(\theta)}.$$

- Intuitively,

$$2LR \approx \sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right)' \frac{1}{n} \frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'} \sqrt{n}\left(\hat{\theta}_{MLE} - \theta_0\right).$$

- Not UMP.

- Not longer χ^2 limit for multivariate inequality test.

- Note that $\frac{\partial \log L(\hat{\theta}_{MLE})}{\partial \theta} = 0$, use a second order Taylor expansion,

$$\begin{aligned} 2LR &= \left(\hat{\theta} - \theta_0\right)' \frac{\partial^2 \log L\left(\theta^*\right)}{\partial \theta \partial \theta'} \left(\hat{\theta} - \theta_0\right) \\ &= \sqrt{n}\left(\hat{\theta} - \theta_0\right)' \frac{1}{n} \frac{\partial^2 \log L\left(\theta^*\right)}{\partial \theta \partial \theta'} \sqrt{n}\left(\hat{\theta} - \theta_0\right) \end{aligned}$$

- Note that

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, H^{-1} S H^{-1}\right) = N\left(0, -H^{-1}\right)$$

The 2nd equality holds if the model is correct.

$$\frac{1}{n} \frac{\partial^2 \log L\left(\theta^*\right)}{\partial \theta \partial \theta'} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f\left(X_i | \theta^*\right)}{\partial \theta \partial \theta'}$$

As $\hat{\theta} \xrightarrow{p} \theta_0$, $\theta^* \xrightarrow{p} \theta_0$,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f\left(X_i | \theta^*\right)}{\partial \theta \partial \theta'} \xrightarrow{p} E \frac{\partial^2 \log f\left(X_i | \theta_0\right)}{\partial \theta \partial \theta'} \equiv H.$$

- Therefore by Slutsky and CMT, when the model is correct, for $Z \sim N\left(0, -H^{-1}\right)$:

$$2LR \xrightarrow{d} -Z'HZ = Z'\Sigma^{-1/2}\Sigma^{-1/2}Z \sim \chi^2_{\mathrm{dim}(\theta)}$$

- If the model is misspecified (but the parameter is still consistent), then for $Z \sim N\left(0, H^{-1}SH^{-1}\right)$,

$$2LR \xrightarrow{d} -Z'HZ$$

This is some kind of "weighted" chi-square, but no longer χ^2. It can be simulated using consistent estimates $\hat{H}$ and $\hat{S}$.

- More generally, for $\hat{\theta}$ the unconstrained MLE and $\bar{\theta}$ the constrained MLE:

$$LR = \frac{\sup_{\theta \in H_0 \cup H_1} L(X|\theta)}{\sup_{\theta \in H_0} L(X|\theta)} = \frac{L\left(X|\hat{\theta}\right)}{L\left(X|\bar{\theta}\right)}.$$

- $H_0 : R\theta = \gamma$, versus, $H_0 : R\theta \neq \gamma$, then

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(X|\theta)$$
$$\bar{\theta} = \arg\max_{\theta \in \Theta} L(X|\theta) \quad \text{subject to} \quad R\theta = \gamma.$$

Find c_α so that

$$P\left(R = \left(X : \frac{L\left(X|\hat{\theta}\right)}{L\left(X|\bar{\theta}\right)} > c_\alpha\right) |H_0\right) = \alpha.$$

Difficult to derive the exact finite sample distribution.

- It can be shown however, that as $n \to \infty$, under the H_0:

$$2LR = 2\left[\log L\left(X_n|\hat{\theta}\right) - \log L\left(X_n|\bar{\theta}\right)\right] \xrightarrow{d} \chi^2_{df=d_R}$$

The degree of freedom is the number of restrictions.

- Even more generally, $H_0 : g(\theta) = 0$, $H_1 : g(\theta) \neq 0$,

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(X|\theta)$$
$$\bar{\theta} = \arg\max_{\theta\in\Theta} L(X|\theta) \quad \text{subject to} \quad g(\theta) = 0.$$

It is still true that

$$2LR = 2\left[\log L\left(X_n|\hat{\theta}\right) - \log L\left(X_n|\bar{\theta}\right)\right] \xrightarrow{d} \chi^2_{df=d_g}$$

where d_g is the number of constraints in $g(\theta)$ as long as the $\frac{\partial g(\theta_0)}{d\theta}$ has full row rank.

- Asymptotic power.
- Under H_A, typically $a_n t(\mathbf{x}) \longrightarrow \infty$ w.p. $\to 1$.
- Under H_A, reject w.p. $\to 1$.
- Asymptotic power is 1.
- These are called consistent tests.
- Local alternatives and local power.
- Local alternative, $H_A : \theta = \theta_0 + \frac{c}{\sqrt{n}}$.

- Example: $H_0 : \mu = \mu_0$, $H_A : \mu > \mu_0$. X_t is i.i.d such that
- $\{X_t \sim N(\mu, \sigma^2)\}$ with σ^2 known.
- It can be shown that $\bar{X}$ is a sufficient statistic that contains all the information about μ
- Under the H_0: $\bar{X} \sim N\left(0, \frac{\sigma^2}{n}\right)$.
- Reject if $\bar{x} > d$: $d = \frac{\sigma z_\alpha}{\sqrt{n}} + \mu_0$ for size α test, since

$$P\left(\frac{\bar{X}}{\sqrt{1/n}} > \frac{d}{\sqrt{1/n}} | H_0\right) = \alpha.$$

- Asymptotic power: for fixed $\mu_1 > \mu_0$,

$$\begin{aligned} P(\bar{x} > d|\mu_1) =& P\left(\sqrt{n}\frac{\bar{x} - \mu_1}{\sigma} \geq z_\alpha + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}\right) \\ =& 1 - \Phi\left(z_\alpha + \frac{\mu_0 - \mu_1}{\sigma}\sqrt{n}\right) \to 1 \end{aligned}$$

$$P(\bar{x} > d|\mu_1) \to \begin{cases} 1 & \text{when } \mu_1 \to \infty \text{ holding } n \text{ fixed.} \\ 1 & \text{when } n \to \infty \text{ holding } \mu > 0 \text{ fixed.} \end{cases}$$

- Draw the shape of the power function when n increases.
- If $n = \infty$, there should be no Type II error. Implications:
- Why fix α? Since Power $= 1 - \Phi\left(Z_{1-\alpha} - \sqrt{n}\mu\right)$, if you really believe in $n \to \infty$, you can choose $\alpha \to 0$, $Z_{1-\alpha} \to \infty$ in such a way that $Z_{1-\alpha} - \sqrt{n}\mu \to -\infty$. Then both size $\to 0$ and Power $(\mu, n) \to 1$ as $n \to \infty$.
- Local power, for $\mu_1 = \mu_0 + \frac{c}{\sqrt{n}}$,

$$\begin{aligned} P(\bar{x} > d|\mu_1) =& P\left(\sqrt{n}\frac{\bar{x} - \mu_1}{\sigma} \geq z_\alpha - \frac{c}{\sigma}\right) \\ \to& 1 - \Phi\left(z_\alpha - \frac{c}{\sigma}\right) > \alpha. \end{aligned}$$

- P-value: probability of rejection if the observed test statistic is used as the critical value.
- Test-statistic: $T = T(X_1, \ldots, X_n)$. Reject if $T > c$.
- Let $F(\cdot)$ be the CDF of T under the null H_0.

$$\text{Pvalue} = 1 - F(T)$$

- Reject if $T > c$, or if Pvalue $< \alpha$.

$$P(\text{Pvalue} < x) = P(1 - F(T) < x) = P(F(T) > 1 - x)$$
$$= P(T > F{-}1(1 - x)) = 1 - F\left(F^{-1}(1 - x)\right) = 1 - (1 - x) = x.$$

- Relation between Testing and Confidence Set
- Confidence Set: a set S such that

$$P(\theta_0 \in S) \begin{cases} = 1-\alpha \\ \to 1-\alpha \end{cases}$$

- Any test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ can be converted into a confidence set.
- Consider a test statistic $T(X, \theta_0)$. Reject if $T(X, \theta_0) > C_{\theta_0,\alpha}$. Let

$$S = \{\text{set of } \theta_0 \text{ that can not be rejected by the above test.}\}.$$

Then

$$P_{\theta_0}(\theta_0 \in S) = P_{\theta_0}(T(X, \theta_0) \leq C_{\theta_0,\alpha}) = 1-\alpha.$$

For example, in the Binary case,

$$\frac{\sqrt{n}\,(\hat{p}-p)}{\sqrt{\hat{p}\,(1-\hat{p})}} \xrightarrow{d} N\,(0,1)$$

$$\frac{\sqrt{n}\,(\hat{p}-p)}{\sqrt{\hat{p}\,(1-\hat{p})}} \xrightarrow{d} N\,(0,1)\,.$$

If we test: $H_0 : p = \bar{p} \quad H_1 : p \neq \bar{p}$. Reject if

$$T_1\,(\hat{p},\bar{p}) = \left|\frac{\sqrt{n}\,(\hat{p}-\bar{p})}{\sqrt{\hat{p}\,(1-\hat{p})}}\right| > Z_{1-\alpha/2} \quad \text{or}$$

$$T_2\,(\hat{p},\bar{p}) = \left|\frac{\sqrt{n}\,(\hat{p}-\bar{p})}{\sqrt{\bar{p}\,(1-\bar{p})}}\right| > Z_{1-\alpha/2}$$

This implies two confidence sets:

$$S_1 = \left\{\bar{p} \quad \text{such that} \quad \left|\frac{\sqrt{n}\,(\hat{p}-\bar{p})}{\sqrt{\hat{p}\,(1-\hat{p})}}\right| < Z_{1-\alpha/2}\right\}$$

$$S_1 = \left\{\bar{p} \quad \text{such that} \quad \left|\frac{\sqrt{n}\,(\hat{p}-\bar{p})}{\sqrt{\bar{p}\,(1-\bar{p})}}\right| < Z_{1-\alpha/2}\right\}$$

- Exact finite sample pivotal test.
- The T-statistic is finite sample pivotal if X_i is a known location scale family.
- Suppose $X_i \sim \frac{1}{\sigma} f\left(\frac{X_i - \mu}{\sigma}\right)$. So that $X_i = \mu + \sigma Z_i$, where $Z_i \sim f(\cdot)$ and $f(\cdot)$ is known. Then

$$T = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}} = \frac{\sqrt{n}\left(\bar{Z}\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)^2}}$$

- This distribution can be simulated as long as you know $f(\cdot)$. It is pivotal since it is free of nuisance parameters.
- In general the T-statistic is asymptotically pivotal.

$$T \xrightarrow{d} N(0,1).$$