Lecture 19 Asymptotic Distributions of LR Test Statistics

Except for a few special cases, distributions of LR test statistics for finite samples are not available. Large sample approximations become a useful general approach.

Let $X_1, X_2, ...$ be iid samples from $f(x|\theta)$, $\theta \in \Theta$. For testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta \backslash \Theta_0$, consider the LR statistic $\log \Lambda(X^n) = \sup_{\theta \in \Theta} L(\theta|X^n) - \sup_{\theta \in \Theta_0} L(\theta|X^n)$ where $L(\theta|X^n) = \log \prod_{i=1}^n f(X_i|\theta)$.

Theorem 1 *Let $\Theta_0 = \{\theta_0\} \in \mathbb{R}$. Under certain regularity conditions similar to those required for Theorem 1 in Lecture 18,*

$$2 \ \log \Lambda(X^n) \xrightarrow{\mathcal{D}} \chi_1^2 \quad \text{under } P_{\theta_0}, \tag{19.1}$$

where χ_k^2 denotes a chi squared distribution with k degrees of freedom.

Proof: Assume MLE $\hat{\theta}$ is the unique solution to the likelihood equation, i.e. $\left[\frac{d}{d\theta} L(\theta|X^n)\right]_{\theta=\hat{\theta}} = 0$. Then

$$2 \ \log \Lambda(X^n) = 2 \ [L(\hat{\theta}|X^n) - L(\theta_0|X^n)] = - \left[\frac{d^2}{d\theta^2} L(\theta|X^n)\right]_{\theta=\theta^*} (\hat{\theta} - \theta_0)^2,$$

where θ^* is between θ_0 and $\hat{\theta}$. Notice the following facts:

(a) $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, [I_{X_1}(\theta_0)]^{-1})$ hence $n \ I_{X_1}(\theta_0) \ (\hat{\theta} - \theta_0)^2 \xrightarrow{\mathcal{D}} \chi_1^2$.

(b) $-\frac{1}{n} \left[\frac{d^2}{d\theta^2} L(\theta|X^n)\right]_{\theta=\theta_0} = \frac{1}{n} \sum_{i=1}^n \left[-\frac{d^2}{d\theta^2} L(\theta|X_i)\right]_{\theta=\theta_0} \xrightarrow{P_{\theta_0}} I_{X_1}(\theta_0)$.

(c) $-\frac{1}{n} \left[\frac{d^2}{d\theta^2} L(\theta|X^n)\right]_{\theta=\theta^*} + \frac{1}{n} \left[\frac{d^2}{d\theta^2} L(\theta|X^n)\right]_{\theta=\theta_0} = -\frac{1}{n} \left[\frac{d^3}{d\theta^3} L(\theta|X^n)\right]_{\theta=\theta'} (\theta^* - \theta_0) \xrightarrow{P_{\theta_0}} 0$, where θ' is between θ_0 and θ^*, and we assume $\left|\frac{d^3}{d\theta^3} L(\theta|x)\right| \leq M(x) \ \forall \, \theta$ with $E_{\theta_0} M(X_1) < \infty$.

(19.1) follows from (a), (b), (c) and the Slutsky's Theorem. *QED*.

Note:

(i) An extension of Theorem 1 to the case with multidimensional parameters and a composite H_0 is usually referred to as Wilks' Theorem: Let Θ be an open subset of $\mathbb{R}^k$, and $\Theta_0 = \{\theta : \ g(\theta) = c\}$, where g is a C^1 map and $c = (c_1, ..., c_r)$. Then $2 \log \Lambda(X^n) \xrightarrow{\mathcal{D}} \chi_r^2$ under the null hypothesis H_0. See the book by Bickel and Doksum for the proof. Denote the dimensionality of Θ (definition?) by $dim\Theta$. Then the degree of freedom for the limiting χ^2 distribution is equal to $dim\Theta - dim\Theta_0 = k - (k - r) = r$. An important feature of Wilks' Theorem is that the limiting χ^2 distribution does not depend on any particular $\theta \in \Theta_0$.

(ii) Suppose $\theta_0 \in \Theta_0$ and $\theta_n = \theta_0 + \delta/\sqrt{n} \neq \Theta_0$ where $\delta \in I\!R^k$. The local power at θ_n can be calculated approximately: $\beta(\theta_n) = P_{\theta_n}(2 \log \Lambda(X^n) > b) \approx P(\chi_r^2(\eta) > b)$ where b is determined by $P(\chi_r^2 > b) = \alpha$ (a significance level of the test), and $\eta = \delta\ I_{X_1}(\theta_0)\ \delta^t$ is the non-centrality factor in the non-central $\chi_r^2(\eta)$ distribution.