STAT-511	**2018 Spring**

Lecture 1: Likelihood Ratio Tests and Beyond.

Instructor: Naveen N. Narisetty   Scribe: Yutong Dai   Last Modified: 2018-05-04

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications, and they may contain factual and/or typographic errors.*

1.1 Likelihood Ratio Tests

1.1.1 Asymptotic Distribution Under the Null

Theorem 1.1 For hypothesis testing problem,

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_A,$$

where both Θ_0, Θ_A are open sets. Then the likelihood ratio test statsistics $\lambda(\mathbf{X})$ statisfies,

Under $H_0, 2\log\lambda(\mathbf{X}) \xrightarrow{D} \chi^2_k$, where $k = dim(\Theta_1) - dim(\Theta_0)$.

Example 1.2 (existense of nuisance parameters)

Refer to Section 16.1.1 - Section 16.1.3 (p203-p207) from John Marden's book.

From examples above, we can see, if the data are from gaussian, we can hopefully find the exact distributions of those test statistics. But it's quite convenient to use its asymptotic null distribution in general cases.

1.2 Score Tests (Book 16.3)

When the null hypothesis is simple, tests based directly on the score function can often be simpler to implement than the LRT, since we do not need to find the MLE under the alternative. Instead, we can approximate the Likelihood Ratio statistic $\lambda(x)$ as the score function at the null hypothesis.

1.2.1 One-sided Tests

Start with $X_1, ..., X_n \stackrel{iid}{\sim} f(x;\theta), \theta \in R$. Consider the one-side test

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta > \theta_0,$$

$\forall \theta_A > \theta_0$, the log-likelihood ratio test takes the form of

$$\sum_{i=1}^{n} \log(\frac{f(x_i;\theta_A)}{f(x_i;\theta_0)}) \triangleq l_n(\theta_A;\mathbf{X}) - l_n(\theta_0;\mathbf{X}) \approx (\theta_A - \theta_0)\nabla_\theta l_n(\theta_0;\mathbf{X}),$$

where we denote $l_n(\theta;\mathbf{X}) = \sum_{i=1}^n \log(f(x_i;\theta)$.

If θ_A is fixed, then $(\theta_A - \theta_0)\nabla_\theta l_n(\theta_0;\mathbf{X}) > c$ is equivalent to $\nabla_\theta l_n(\theta_0;\mathbf{X}) > c^*$. Therefore, The statistic $\nabla_\theta l_n(\theta_0;\mathbf{X})$ is approximately equivalent to the likelihood ratio statistic, which is the best statistic, when θ_A is close to θ_0.

Since we are in iid case, $\nabla_\theta l_n(\theta_0;\mathbf{X}) = \sum_{i=1}^n \nabla_\theta \log f(x_i;\theta)|_{\theta_0}$, the score for one observation.

Under the null hypothesis, we have

$$E_{\theta_0}(\nabla_\theta f(x_i;\theta)|_{\theta_0}) = 0, \quad Var_{\theta_0}(\nabla_\theta f(x_i;\theta)|_{\theta_0}) = I_1(\theta_0).$$

By the central limit theorem, we have

$$\sqrt{n}(\frac{1}{n}\nabla_\theta l_n(\theta_0;\mathbf{X})) \xrightarrow{D} N(0, I_1(\theta_0)).$$

1.2.2 Many-sided Tests

Suppose $X_1, ..., X_n \overset{iid}{\sim} f(x;\theta), \theta \in \Theta \subset R^k$. Consider the test

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \in \Theta - \{\theta_0\},$$

Similarly, under the H_0, we have

$$\frac{\nabla_\theta l_n(\theta_0;\mathbf{X})}{\sqrt{n}} \xrightarrow{D} N(0, I_1(\theta_0)),$$

where

$$I_1(\theta_0) = Var(\nabla_\theta l_1(\theta_0;x_i)) = E_{\theta_0}\begin{pmatrix} \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,1}^2} & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,1}\partial\theta_{0,2}} & \cdots & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,1}\partial\theta_{0,k}} \\ \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,2}\partial\theta_{0,1}} & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,2}^2} & \cdots & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,2}\partial\theta_{0,k}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,k}\partial\theta_{0,1}} & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,k}\partial\theta_{0,2}} & \cdots & \frac{\partial^2 l_1(\theta_0;x_i)}{\partial\theta_{0,k}^2} \end{pmatrix}.$$

So

$$S_n^2 = \frac{1}{n}\nabla_\theta^T l_n(\theta_0;\mathbf{X}) I_1^{-1}(\theta_0)\nabla_\theta l_n(\theta_0;\mathbf{X}) \xrightarrow{D} \chi_k^2.$$

The approximate size α test rejects the null if $S_n^2 > \chi_k^2(1-\alpha)$.