557: MATHEMATICAL STATISTICS II
LARGE SAMPLE AND ASYMPTOTIC RESULTS IN HYPOTHESIS TESTING

Recall that the **Likelihood Ratio Test (LRT)** statistic for testing H_0 against H_1

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned} \tag{1}$$

is based on the statistic

$$\lambda_{\mathbf{X}}(\mathbf{x}) = \frac{\sup\limits_{\theta \in \Theta_0} f_{\mathbf{X}}(\mathbf{x};\theta)}{\sup\limits_{\theta \in \Theta_0 \cup \Theta_1} f_{\mathbf{X}}(\mathbf{x};\theta)} = \frac{\mathscr{L}_n(\widehat{\boldsymbol{\theta}}_{n0})}{\mathscr{L}_n(\widehat{\boldsymbol{\theta}}_n)} \tag{2}$$

say, where H_0 is **rejected** if $\lambda_{\mathbf{X}}(\mathbf{x})$ is **small enough**, that is, $\lambda_{\mathbf{X}}(\mathbf{x}) \le c$ for some c to be defined.

Theorem Consider using the likelihood ratio statistic $\lambda_{\mathbf{X}}(\mathbf{X})$ for testing

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned} \tag{3}$$

for a $k \times 1$ parameter θ. Then under regularity assumptions, the maximum likelihood estimator for θ under H_1, $\widehat{\theta}_n$, is consistent and asymptotically normally distributed, and as $n \longrightarrow \infty$ under H_0,

$$-2(\ell_n(\theta_0) - \ell_n(\widehat{\theta}_n)) \xrightarrow{d} Q \sim \chi^2_k.$$

For the behaviour under H_1, suppose that, the true data generating model has $\theta = \theta_0^*$. Then, under regularity conditions

$$-\frac{2}{n}(\ell_n(\theta_0)-\ell_n(\widehat{\theta}_n)) = \frac{2}{n}(\ell_n(\theta_0^*)-\ell_n(\theta_0))-\frac{2}{n}(\ell_n(\theta_0^*)-\ell_n(\widehat{\theta}_n)) = \frac{2}{n}\sum_{i=1}^n \log\frac{f_X(X_i;\theta_0^*)}{f_X(X_i;\theta_0)}-\frac{2}{n}(\ell_n(\theta_0^*)-\ell_n(\widehat{\theta}_n))$$

But, as $n \longrightarrow \infty$

$$\frac{2}{n}\sum_{i=1}^n \log\frac{f_X(X_i;\theta_0^*)}{f_X(X_i;\theta_0)} \xrightarrow{p} 2K(\theta_0^*,\theta_0) \qquad\qquad \frac{2}{n}(\ell_n(\theta_0^*)-\ell_n(\widehat{\theta}_n)) \xrightarrow{p} 0$$

as $\widehat{\theta}_n \xrightarrow{p} \theta_0^*$, so

$$-\frac{2}{n}(\ell_n(\theta_0)-\ell_n(\widehat{\theta}_n)) \xrightarrow{p} 2K(\theta_0^*,\theta_0) > 0$$

Theorem Consider testing the hypothesis in (1) using the likelihood ratio statistic (2). Suppose H_0 specifies a model that constrains $k-k_1$ parameters, and $H_0 \cup H_1$ and constrains only a subset of $k-k_2$ parameters, with $k_2 > k_1$; that is, H_0 imposes a further $k_2 - k_1$ constraints compared to those imposed by H_1. Then, under regularity conditions, as $n \longrightarrow \infty$, under H_0

$$-2\log\lambda_{\mathbf{X}}(\mathbf{X}) = -2(\ell_n(\widehat{\theta}_{n0}) - \ell_n(\widehat{\theta}_n)) \xrightarrow{d} Q \sim \chi^2_{k_2-k_1}$$

Such hypotheses can often be specified in the form of parameter $\psi = (\theta, \phi)$,

$$\begin{aligned} H_0 &: \theta = \theta_0, \qquad \phi \text{ unspecified} \\ H_1 &: \theta \neq \theta_0, \qquad \phi \text{ unspecified} \end{aligned}$$

that is, H_0 places constraints on one component, θ, of ψ, but leaves the other component, ϕ, unconstrained.

OTHER ASYMPTOTIC TESTS: THE WALD AND SCORE STATISTICS

The **Wald** and **Rao/Score** test statistics derived from a sample of size n, W_n and R_n, for testing (3) in the case of a vector-valued parameter are constructed as follows:

- **Wald Test :** The **Wald Statistic**, W_n, is defined by

$$W_n = n(\widehat{\theta}_n - \theta_0)^\top \widehat{I}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_0) \tag{4}$$

  where $\widetilde{\boldsymbol{\theta}}_n$ is a solution to the likelihood equations, $\widehat{I}_n$ is the observed information.

- **Score Test :** Let

$$Z_n \equiv Z_n(\theta_0) = \frac{1}{\sqrt{n}}\,\dot{\ell}_n(\theta_0).$$

  Then the (Rao) Score Test Statistic, R_n, is defined by

$$R_n = Z_n^\top \mathcal{I}_{\theta_0}(\theta_0)^{-1} Z_n \tag{5}$$

  where $\mathcal{I}_{\theta_0}(\theta_0)$ can be replaced by the observed information $\widehat{I}_n(\theta_0)$ if necessary.

In the one parameter case, the statistics can be expressed as

$$W_n = -(\widehat{\theta}_n - \theta_0)^2 \ddot{\ell}_n(\widehat{\theta}_n) \qquad\qquad R_n = -\left\{\dot{\ell}_n(\theta_0)\right\}^2 \left\{\ddot{\ell}_n(\theta_0)\right\}^{-1}$$

(a) Under the **null hypothesis**

- **Wald Test :** For the Wald test, as

$$D_n = \sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Z \sim \text{Normal}(\mathbf{0}_k, \mathcal{I}_{\theta_0}(\theta_0)^{-1})$$

  it follows that

$$W_n = n(\widehat{\theta}_n - \theta_0)^\top \widehat{I}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_0) = D_n^\top \widehat{I}_n(\widehat{\theta}_n) D_n \xrightarrow{d} Z^\top \mathcal{I}_{\theta_0}(\theta_0) Z \sim \chi^2_k$$

- **Score Test :** For the Score test,

$$Z_n = \frac{1}{\sqrt{n}}\,\dot{\ell}_n(\theta_0) = \sqrt{n}\left(\frac{1}{n}\,\dot{\ell}_n(\theta_0)\right) \xrightarrow{d} Z \sim \text{Normal}(\mathbf{0}_k, \mathcal{I}_{\theta_0}(\theta_0))$$

  and hence for the Score Test Statistic,

$$R_n = Z_n^\top \{\mathcal{I}_{\theta_0}(\theta_0)\}^{-1} Z_n \xrightarrow{d} Z^\top \mathcal{I}_{\theta_0}(\theta_0) Z \sim \chi^2_k$$

(b) If the null hypothesis is **not true**, let θ_0^* denote the true value of the parameter.

- **Wald Test :** For the Wald test,

$$\frac{1}{n} W_n = (\widehat{\theta}_n - \theta_0)^\top \widehat{I}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_0) \xrightarrow{p} (\theta_0^* - \theta_0)^\top \mathcal{I}_{\theta_0^*}(\theta_0^*)(\theta_0^* - \theta_0) > 0$$

- **Score Test :** For the Score test, from above

$$\frac{1}{\sqrt{n}} Z_n = \frac{1}{n}\,\dot{\ell}_n(\theta_0) \xrightarrow{p} \mathbb{E}_X[\dot{\ell}_n(\theta_0)] = \mu(\theta_0^*, \theta_0)$$

  say, and hence for the Score Test Statistic,

$$\frac{1}{n} R_n = \frac{1}{n} Z_n^\top \{\mathcal{I}_{\theta_0}(\theta_0)\}^{-1} Z_n \xrightarrow{p} \mu(\theta_0^*, \theta_0)^\top \{\mathcal{I}_{\theta_0}(\theta_0)\}^{-1} \mu(\theta_0^*, \theta_0) > 0$$

Composite Hypotheses
Consider testing the hypothesis for parameter $\psi = (\theta, \phi)$ where θ is $k_1 \times 1$, and ϕ is $(k - k_1) \times 1$

$$\begin{aligned} H_0 &: \theta = \theta_0, \qquad \phi \text{ unspecified} \\ H_1 &: \theta \neq \theta_0, \qquad \phi \text{ unspecified} \end{aligned}$$

that is, H_0 places constraints on one component of ψ, but leaves the other unspecified.

Let $\widehat{\psi}_{n0} = (\theta_0, \widehat{\phi}_{n0})^\top$ and $\widehat{\psi}_{n1} = (\widehat{\theta}_{n1}, \widehat{\phi}_{n1})^\top$ denote MLEs under H_0 and H_1 respectively. Let

$$\mathcal{I}_\psi(\psi) = \begin{bmatrix} \mathcal{I}_{00}(\psi) & \mathcal{I}_{01}(\psi) \\ \mathcal{I}_{10}(\psi) & \mathcal{I}_{11}(\psi) \end{bmatrix}$$

denotes the Fisher information for θ, with blocks $\mathcal{I}_{00}(\psi)$ $(k_1 \times k_1)$, $\mathcal{I}_{01}(\psi)$ $(k_1 \times (k - k_1))$ etc. Let

$$\{\mathcal{I}_\psi(\psi)\}^{-1} = \begin{bmatrix} \widetilde{\mathcal{I}}_{00}(\psi) & \widetilde{\mathcal{I}}_{01}(\psi) \\ \widetilde{\mathcal{I}}_{10}(\psi) & \widetilde{\mathcal{I}}_{10}(\psi) \end{bmatrix}$$

Then for the hypotheses above the Wald and Score tests are constructed as follows:

- **Wald Test :** The **Wald Statistic**, W_n, is defined by

$$W_n = n(\widehat{\theta}_{n1} - \theta_0)^\top \widehat{I}_{n00.1}(\widehat{\psi}_{n1})(\widehat{\theta}_{n1} - \theta_0) \tag{6}$$

  where $\widehat{I}_n$ is the observed information, and $\widehat{I}_{n00.1}$ is the upper $k_1 \times k_1$ block of the inverse of $\widehat{I}_n(\widehat{\psi}_{n1})$. It can be shown that if

$$\widehat{I}_n(\widehat{\psi}_{n1}) = \begin{bmatrix} \widehat{I}_{n00}(\widehat{\psi}_{n1}) & \widehat{I}_{n01}(\widehat{\psi}_{n1}) \\ \widehat{I}_{n10}(\widehat{\psi}_{n1}) & \widehat{I}_{n11}(\widehat{\psi}_{n1}) \end{bmatrix}$$

  with inverse

$$\widetilde{I}_n(\widehat{\psi}_{n1}) = \begin{bmatrix} \widetilde{I}_{n00}(\widehat{\psi}_{n1}) & \widetilde{I}_{n01}(\widehat{\psi}_{n1}) \\ \widetilde{I}_{n10}(\widehat{\psi}_{n1}) & \widetilde{I}_{n11}(\widehat{\psi}_{n1}) \end{bmatrix}$$

  then

$$\widehat{I}_{n00.1}(\widehat{\theta}_{n0}) = \left(\widetilde{I}_{n00}(\widehat{\psi}_{n1}) - \widetilde{I}_{n01}(\widehat{\psi}_{n1})\{\widetilde{I}_{n11}(\widehat{\psi}_{n1})\}^{-1}\widetilde{I}_{n10}(\widehat{\psi}_{n1})\right)^{-1}$$

- **Score Test :** Let

$$Z_n \equiv Z_n(\widehat{\psi}_{n0}) = \frac{1}{\sqrt{n}}\,\dot{\ell}_n(\widehat{\psi}_{n0}).$$

  Then the (Rao) Score Test Statistic, R_n, is defined by

$$R_n = Z_n^\top \{\mathcal{I}_{\widehat{\psi}_{n0}}(\widehat{\psi}_{n0})\}^{-1} Z_n \tag{7}$$

  where $\mathcal{I}_\psi(\psi)$ can be replaced by the observed information $\widehat{I}_n(\widehat{\psi}_{n0})$ if the Fisher information is not available.

In both cases, under H_0, the statistics converge in distribution to a Chi-squared distribution,

$$W_n \xrightarrow{d} \chi^2_{k_1} \qquad\qquad R_n \xrightarrow{d} \chi^2_{k_1}$$