557: MATHEMATICAL STATISTICS II
LARGE SAMPLE AND ASYMPTOTIC RESULTS - III

Behaviour of the Likelihood Ratio Test Statistic

In the test of

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

using the likelihood ratio test, suppose that, in fact $\theta = \theta_T$. Then, under conditions A0-A4,

$$\begin{aligned} -\frac{2}{n}(l_n(\theta_0) - l_n(\widehat{\theta}_n)) &= \frac{2}{n}(l_n(\theta_T) - l_n(\theta_0)) - \frac{2}{n}(l_n(\theta_T) - l_n(\widehat{\theta}_n)) \\ &= \frac{2}{n}\sum_{i=1}^{n}\log\frac{f_{X|\theta}(X_i|\theta_T)}{f_{X|\theta}(X_i|\theta_0)} - \frac{2}{n}(l_n(\theta_T) - l_n(\widehat{\theta}_n)) \end{aligned}$$

But, as $n \longrightarrow \infty$

$$\frac{2}{n}\sum_{i=1}^{n}\log\frac{f_{X|\theta}(X_i|\theta_T)}{f_{X|\theta}(X_i|\theta_0)} \xrightarrow{p} 2K(\theta_T, \theta_0) \qquad \frac{2}{n}(l_n(\theta_T) - l_n(\widehat{\theta}_n)) \xrightarrow{p} 0$$

as $\widehat{\theta}_n \xrightarrow{p} \theta_T$, so

$$-\frac{2}{n}(l_n(\theta_0) - l_n(\widehat{\theta}_n)) \xrightarrow{p} 2K(\theta_T, \theta_0)$$

d-dimensional Parameters

Consider using the likelihood ratio statistic for testing H_0 versus H_1 if $\underset{\sim}{\theta}$ is $d-$dimensional parameter. Under assumptions A0-A4, as $n \longrightarrow \infty$ under H_0,

$$-2\log\lambda_{\underset{\sim}{X}}(\underset{\sim}{X}) = -2(l_n(\underset{\sim}{\theta}_0) - l_n(\widehat{\underset{\sim}{\theta}}_n)) \xrightarrow{d} Q \sim \chi^2_d$$

This follows using identical methods to the $d = 1$ case. A second-order Taylor expansion to the log-likelihood around the MLE, $\widehat{\underset{\sim}{\theta}}_n$

$$l_n(\underset{\sim}{\theta}) = l_n(\widehat{\underset{\sim}{\theta}}_n) + (\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n)^{\mathsf{T}}\dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n) + \frac{1}{2}(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n)^{\mathsf{T}}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n)(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n) + o_P(1)$$

As $\widehat{\theta}_n$ is the maximum likelihood estimate

$$\dot{l}_n(\widehat{\theta}_n) = 0$$

and therefore on rearrangement, evaluating at $\underset{\sim}{\theta} = \underset{\sim}{\theta}_0$,

$$-2(l_n(\theta_0) - l_n(\widehat{\theta}_n)) = -(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n)^{\mathsf{T}}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n)(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n) + o_P(1)$$

But, by previous results

$$(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n)^{\mathsf{T}}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n)(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n) = \sqrt{n}(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n)^{\mathsf{T}}\left(-\frac{1}{n}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n)\right)\sqrt{n}(\underset{\sim}{\theta} - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{d} \underset{\sim}{Z}^{\mathsf{T}}\mathcal{I}(\underset{\sim}{\theta}_0)\underset{\sim}{Z}$$

as

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} \underset{\sim}{Z} \sim \text{Normal}(\underset{\sim}{0}, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}) \qquad -\frac{1}{n}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \mathcal{I}(\underset{\sim}{\theta}_0)$$

Hence

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

Theorem Consider testing the hypothesis

$$\begin{aligned} H_0 &: \underset{\sim}{\theta} \in \Theta_0 \\ H_1 &: \underset{\sim}{\theta} \in \Theta_1 \end{aligned}$$

using the likelihood ratio statistic

$$\lambda_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{\sup_{\underset{\sim}{\theta} \in \Theta_0} f_{\underset{\sim}{X}|\underset{\sim}{\theta}}(\underset{\sim}{x}|\underset{\sim}{\theta})}{\sup_{\underset{\sim}{\theta} \in \Theta} f_{\underset{\sim}{X}|\underset{\sim}{\theta}}(\underset{\sim}{x}|\underset{\sim}{\theta})} = \frac{L(\widehat{\underset{\sim}{\theta}}_{n0} \mid \underset{\sim}{x})}{L(\widehat{\underset{\sim}{\theta}}_{n} \mid \underset{\sim}{x})}$$

say, where $\Theta \equiv \Theta_0 \cup \Theta_1$, and where H_0 specifies a model with k_1 free parameters (parameters not determined by the hypothesis), and H_1 and specifies a model with k_2 free parameters, with $k_2 > k_1$. Then, under assumptions A0-A4, as $n \longrightarrow \infty$, under H_0

$$-2 \log \lambda_{\underset{\sim}{X}}(\underset{\sim}{X}) = -2(l_n(\widehat{\underset{\sim}{\theta}}_{n0}) - l_n(\widehat{\underset{\sim}{\theta}}_{n})) \stackrel{d}{\longrightarrow} Q \sim \chi^2_{k_2-k_1}$$

Note: Such hypotheses can often be specified in the form

$$\begin{aligned} H_0 &: \underset{\sim}{\theta} = (\underset{\sim}{\theta}_0, \underset{\sim}{\theta}_1), \qquad \underset{\sim}{\theta}_1 \text{ unspecified} \\ H_1 &: \underset{\sim}{\theta} \neq (\underset{\sim}{\theta}_0, \underset{\sim}{\theta}_1), \qquad \underset{\sim}{\theta}_1 \text{ unspecified} \end{aligned}$$

that is, H_0 places constraints on one component $\underset{\sim}{\theta}$, but leaves the other unspecified.

Other Asymptotic Tests

The Wald and Rao/Score test statistics derived from a sample of size n, W_n and R_n, for testing

$$\begin{aligned} H_0 &: \underset{\sim}{\theta} = \underset{\sim}{\theta}_0 \\ H_1 &: \underset{\sim}{\theta} \neq \underset{\sim}{\theta}_0 \end{aligned}$$

are constructed as follows:

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

$$W_n = n(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0)^{\mathsf{T}} \widehat{I}_n(\widetilde{\underset{\sim}{\theta}}_n)(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \tag{1}$$

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

- **Score Test :** Let

$$\underset{\sim}{Z}_n \equiv \underset{\sim}{Z}_n\left(\underset{\sim}{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underset{\sim}{\dot{l}}_n(\underset{\sim}{\theta}_0).$$

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

$$R_n = \underset{\sim}{Z}_n^{\mathsf{T}} \mathcal{I}\left(\underset{\sim}{\theta}_0\right)^{-1} \underset{\sim}{Z}_n \tag{2}$$

  where $\mathcal{I}\left(\underset{\sim}{\theta}_0\right)$ can be replaced by the observed information $\widehat{I}_n(\underset{\sim}{\theta}_0)$ if necessary.

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

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

Example: Poisson. For $\theta > 0$, if $s_n = \sum_{i=1}^{n} x_i$, then

$$\begin{aligned} l_n(\theta) &= -n\theta + s_n \log\theta - \sum_{i=1}^{n} \log x_i! \\ \dot{l}_n(\theta) &= -n + s_n/\theta \\ \ddot{l}_n(\theta) &= -s_n/\theta^2 \end{aligned}$$

and hence the MLE, from $\dot{l}_n(\widehat{\theta}_n) = 0$, is $\widehat{\theta}_n = s_n/n = \overline{x}$, with estimator $S_n/n = \overline{X}$. Thus

- **Wald Statistic :** using the formula above

$$W_n = -(\widetilde{\theta}_n - \theta_0)^2 \ddot{l}_n(\widetilde{\theta}_n) = -(\overline{X} - \theta_0)^2 \left(-S_n/(\overline{X})^2\right) = n(\overline{X} - \theta_0)^2/\overline{X}.$$

- **Rao Statistic :** In this case, we can compute the Fisher Information $\mathcal{I}(\theta_0)$ exactly - we have

$$\mathcal{I}(\theta_0) = \mathrm{E}_{f_{X|\theta}}\left[-\Psi(\theta_0, X)\right] = \mathrm{E}_{f_{X|\theta}}\left[X/\theta_0^2\right] = \frac{1}{\theta_0^2}\mathrm{E}_{f_{X|\theta}}[X] = \frac{\theta_0}{\theta_0^2} = \frac{1}{\theta_0}$$

so

$$R_n = \frac{\{Z_n(\theta_0)\}^2}{I(\theta_0)} = \frac{\left(\frac{1}{\sqrt{n}}\left(S_n/\theta_0 - n\right)^2\right)}{1/\theta_0} = \frac{\theta_0}{n}\left(S_n/\theta_0 - n\right)^2 = \frac{n(\overline{X} - \theta_0)^2}{\theta_0}$$

However, using the observed information,

$$R_n = -\left\{\dot{l}_n(\theta_0)\right\}^2 \left\{\ddot{l}_n(\theta_0)\right\}^{-1} = \frac{-\left(S_n/\theta_0 - n\right)^2}{-S_n/\theta_0^2} = \frac{(S_n - n\theta_0)^2}{S_n} = \frac{n(\overline{X} - \theta_0)^2}{\overline{X}}$$

that is, identical to Wald.

- **Likelihood Ratio Statistic**:

$$\lambda_X(\underset{\sim}{x}) = \frac{L_n(\widehat{\theta}_n)}{L_n(\theta_0)} = \frac{e^{-n\widehat{\theta}_n}\widehat{\theta}_n^{S_n}}{e^{-n\theta_0}\theta_0^{S_n}} = \exp\left\{-n(\widehat{\theta}_n - \theta_0) + S_n(\log\widehat{\theta}_n - \log\theta_0)\right\}$$

or equivalently

$$2\log\lambda_X(\underset{\sim}{x}) = -2n(\widehat{\theta}_n - \theta_0) + 2S_n(\log\widehat{\theta}_n - \log\theta_0)$$

Example: Normal. Under the normal model, the likelihood is

$$L_n(\mu, \sigma) = f_{X|\mu,\sigma}(x|\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

and thus, in terms of the random variables, for general X,

$$l(X|\theta) = \log f_{X|\mu,\sigma}(X|\mu, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X - \mu)^2$$

and, for μ

$$\frac{\partial}{\partial\mu} l(X|\theta) = \frac{1}{\sigma^2}(X - \mu) \qquad \frac{\partial^2}{\partial\mu^2}\left\{l(X|\theta)\right\} = -\frac{1}{\sigma^2}$$

whereas for σ^2

$$\frac{\partial}{\partial\sigma^2}\left\{l(X|\theta)\right\} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(X-\mu)^2 \qquad \frac{\partial^2}{\partial(\sigma^2)^2}\left\{l(X|\theta)\right\} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2$$

and

$$\frac{\partial^2}{\partial\mu\partial\sigma^2}\left\{l(X|\theta)\right\} = -\frac{1}{\sigma^4}(X-\mu)$$

(here taking σ^2 as the variable with which we differentiating with respect to). Now

$$\mathrm{E}_{f_{X|\mu,\sigma}}\left[(X-\mu)\right] = 0 \qquad \mathrm{E}_{f_{X|\mu,\sigma}}\left[(X-\mu)^2\right] = \sigma^2$$

we have for the Fisher Information for (μ, σ^2) from a single data point as

$$\begin{aligned}
\mathcal{I}(\mu,\sigma^2) &= -\begin{bmatrix} \mathrm{E}\left[-1/\sigma^2\right] & \mathrm{E}\left[-(X-\mu)/\sigma^4\right] \\ \mathrm{E}\left[-(X_1-\mu)/\sigma^4\right] & \mathrm{E}\left[1/(2\sigma^4)-(X-\mu)^2/\sigma^6\right] \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4} \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix}
\end{aligned}$$

To test

$$\begin{aligned}
H_0 &: (\mu,\sigma) = \underset{\sim}{\theta}_0 = \left(0, \sigma_0^2\right) \\
H_1 &: (\mu,\sigma) \neq \underset{\sim}{\theta}_0.
\end{aligned}$$

such W_n and R_n can be constructed. Under H_0, the μ and σ^2 are completely specified, whereas under H_1, the MLEs of μ and σ^2 are

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}\,)^2.$$

and therefore

$$\widehat{I}_n(\underset{\sim}{\widetilde{\theta}}_n) = \begin{bmatrix} \dfrac{1}{\widehat{\sigma}^2} & 0 \\ 0 & \dfrac{1}{2\widehat{\sigma}^4} \end{bmatrix}$$

Hence the Wald Statistic is

$$\begin{aligned}
W_n &= n(\widetilde{\theta}_n - \theta_0)^{\mathsf{T}}\widehat{I}_n(\underset{\sim}{\widetilde{\theta}}_n)(\widetilde{\theta}_n - \theta_0) \\
&= \begin{bmatrix} \sqrt{n}(\overline{X}-0) \\ \sqrt{n}(S^2-\sigma_0^2) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \dfrac{1}{S^2} & 0 \\ 0 & \dfrac{1}{2S^4} \end{bmatrix} \begin{bmatrix} \sqrt{n}(\overline{X}-0) \\ \sqrt{n}(S^2-\sigma_0^2) \end{bmatrix} \\
&= \frac{n(\overline{X})^2}{S^2} + \frac{n(S^2-\sigma_0^2)^2}{2S^4}
\end{aligned}$$

Asymptotic Properties of the Wald and Score Statistics

(a) Under the **null hypothesis**

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

$$\underset{\sim}{D}_n = \sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \xrightarrow{d} \underset{\sim}{Z} \sim \text{Normal}(\underset{\sim}{0}, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})$$

it follows that

$$W_n = n(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0)^\top \widehat{I}_n(\widetilde{\underset{\sim}{\theta}}_n)(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) = \underset{\sim}{D}_n^\top \widehat{I}_n(\widetilde{\underset{\sim}{\theta}}_n)\underset{\sim}{D}_n \xrightarrow{d} \underset{\sim}{Z}^\top \mathcal{I}(\underset{\sim}{\theta}_0)\underset{\sim}{Z} \sim \chi^2_d$$

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

$$\underset{\sim}{Z}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \dot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_0)\right) \xrightarrow{d} \underset{\sim}{Z} \sim \text{Normal}(\underset{\sim}{0}, \mathcal{I}(\underset{\sim}{\theta}_0))$$

and hence for the Score Test Statistic,

$$R_n = \underset{\sim}{Z}_n^\top \mathcal{I}\left(\underset{\sim}{\theta}_0\right)^{-1} \underset{\sim}{Z}_n \xrightarrow{d} \underset{\sim}{Z}^\top \mathcal{I}(\underset{\sim}{\theta}_0)\underset{\sim}{Z} \sim \chi^2_d$$

(b) If the null hypothesis is **not true**, let $\underset{\sim}{\theta}_T$ denote the true value of the parameter.

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

$$\frac{1}{n}W_n = (\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0)^\top \widehat{I}_n(\widetilde{\underset{\sim}{\theta}}_n)(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \xrightarrow{p} (\underset{\sim}{\theta}_T - \underset{\sim}{\theta}_0)^\top \mathcal{I}(\underset{\sim}{\theta}_T)(\underset{\sim}{\theta}_T - \underset{\sim}{\theta}_0) > 0$$

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

$$\frac{1}{\sqrt{n}}\underset{\sim}{Z}_n = \frac{1}{n}\sum_{i=1}^n \dot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_0) \xrightarrow{p} \mathrm{E}_{f_{X|\underset{\sim}{\theta}}}[\dot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_0)] = \underset{\sim}{\mu}(\underset{\sim}{\theta}_T, \underset{\sim}{\theta}_0)$$

say, and hence for the Score Test Statistic,

$$\frac{1}{n}R_n = \frac{1}{n}\underset{\sim}{Z}_n^\top \mathcal{I}\left(\underset{\sim}{\theta}_0\right)^{-1} \underset{\sim}{Z}_n \xrightarrow{p} \underset{\sim}{\mu}(\underset{\sim}{\theta}_T, \underset{\sim}{\theta}_0)^\top \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}\underset{\sim}{\mu}(\underset{\sim}{\theta}_T, \underset{\sim}{\theta}_0) > 0$$

Composite Hypotheses

Consider testing the hypothesis

$$\begin{aligned} H_0 &: \underset{\sim}{\theta} \in \Theta_0 \\ H_1 &: \underset{\sim}{\theta} \in \Theta_1 \end{aligned}$$

where

$$\begin{aligned} H_0 &: \underset{\sim}{\theta} = (\underset{\sim}{\theta}_0, \underset{\sim}{\theta}_1), \qquad \underset{\sim}{\theta}_1 \text{ unspecified} \\ H_1 &: \underset{\sim}{\theta} \neq (\underset{\sim}{\theta}_0, \underset{\sim}{\theta}_1), \qquad \underset{\sim}{\theta}_1 \text{ unspecified} \end{aligned}$$

where $\underset{\sim}{\theta}_0$ is $k_1 \times 1$, and $\underset{\sim}{\theta}_1$ is $(d-k_1)\times 1$, that is, H_0 places constraints on one component $\underset{\sim}{\theta}$, but leaves the other unspecified.

Let $\widetilde{\underset{\sim}{\theta}}_{n0} = (\underset{\sim}{\theta}_0, \widetilde{\underset{\sim}{\theta}}_{n01})^\top$ and $\widetilde{\underset{\sim}{\theta}}_{n1} = (\widetilde{\underset{\sim}{\theta}}_{n10}, \widetilde{\underset{\sim}{\theta}}_{n11})^\top$ denote consistent estimators (possibly MLEs) under H_0 and H_1 respectively.

Sppose that

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

denotes the Fisher information for $\underset{\sim}{\theta}$, with blocks $\mathcal{I}_{00}(\underset{\sim}{\theta})$ $(k_1 \times k_1)$, $\mathcal{I}_{01}(\underset{\sim}{\theta})$ $(k_1 \times (d-k_1))$ etc. Let

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

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

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

$$W_n = n(\widetilde{\underset{\sim}{\theta}}_{n10} - \underset{\sim}{\theta}_0)^\top \widehat{I}_{n00.1}(\widetilde{\underset{\sim}{\theta}}_{n0})(\widetilde{\underset{\sim}{\theta}}_{n10} - \underset{\sim}{\theta}_0) \tag{3}$$

  where $\widetilde{\underset{\sim}{\theta}}_n$ is a solution to the likelihood equations, $\widehat{I}_n$ is the observed information, and $\widehat{I}_{n00.1}$ is the upper $k_1 \times k_1$ block of the inverse if $\widehat{I}_n$. It can be shown that if

$$\widehat{I}_n(\underset{\sim}{\theta}) = \begin{bmatrix} \widehat{I}_{n00}(\underset{\sim}{\theta}) & \widehat{I}_{n01}(\underset{\sim}{\theta}) \\ \widehat{I}_{n10}(\underset{\sim}{\theta}) & \widehat{I}_{n11}(\underset{\sim}{\theta}) \end{bmatrix}$$

  then

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

- **Score Test :** Let

$$\underset{\sim}{Z}_n \equiv \underset{\sim}{Z}_n\left(\underset{\sim}{\theta}_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underset{\sim}{\dot{l}}_n(\underset{\sim}{\theta}_0).$$

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

$$R_n = \underset{\sim}{Z}_n^\top(\widetilde{\underset{\sim}{\theta}}_{n0})\mathcal{I}\left(\widetilde{\underset{\sim}{\theta}}_{n0}\right)^{-1} \underset{\sim}{Z}_n(\widetilde{\underset{\sim}{\theta}}_{n0}) \tag{4}$$

  where $\mathcal{I}\left(\underset{\sim}{\theta}_0\right)$ can be replaced by the observed information $\widehat{I}_n(\underset{\sim}{\theta}_0)$ if necessary.

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

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