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

Observed Information
The asymptotic distribution of the maximum likelihood estimator is given by previous results as

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

For practical purposes, it is often necessary to produce an estimator of $\mathcal{I}(\underset{\sim}{\theta}_0)$. The *observed information*, $\widehat{I}_n = \widehat{I}_n(\underset{\sim}{X})$, can be used to approximate $\mathcal{I}(\underset{\sim}{\theta}_0)$, where any of the definitions

$$\widehat{I}_n(\widetilde{\underset{\sim}{\theta}}_n) = \begin{cases} \mathcal{I}(\widetilde{\underset{\sim}{\theta}}_n) \\ \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \underset{\sim}{S}(X_i, \widetilde{\underset{\sim}{\theta}}_n)\underset{\sim}{S}(X_i, \widetilde{\underset{\sim}{\theta}}_n)^{\mathsf{T}} \\ \\ -\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \Psi(\widetilde{\underset{\sim}{\theta}}_n, X_i) \end{cases}$$

may be used, where $\widetilde{\underset{\sim}{\theta}}_n$ is a consistent estimator of $\underset{\sim}{\theta}_0$. If, as in some hypothesis testing examples, $\underset{\sim}{\theta}_0$ is known (or hypothesized) it replaces $\widetilde{\underset{\sim}{\theta}}_n$ in the calculation.

Improving Inefficient Estimators: The One-Step Estimator

Objective : to produce a consistent estimator with asymptotic variance equal to the inverse Fisher information $\mathcal{I}(\underset{\sim}{\theta}_0)^{-1}$, as this is the best possible variance we can achieve for consistent estimators.

Suppose that $\widehat{\underset{\sim}{\theta}}^{(0)}$ is a (consistent) estimator of $\underset{\sim}{\theta}$ with asymptotic variance $\Sigma^{(0)}$ where $\Sigma^{(0)} - \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}$ is positive definite, so that

$$\Sigma^{(0)} \geq \mathcal{I}(\underset{\sim}{\theta}_0)^{-1} \qquad \text{or} \qquad x^{\mathsf{T}}(\Sigma^{(0)} - \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})x > 0 \qquad \forall\, x \in R^d$$

so that $\widehat{\underset{\sim}{\theta}}^{(0)}$ is inefficient. This estimator can be improved by two iterative procedures that each define a sequence of estimators that are consistent, and will have increasingly better properties.

- **Newton's Method** For $k = 0, 1, \ldots$, define

$$\widehat{\underset{\sim}{\theta}}^{(k+1)} = \widehat{\underset{\sim}{\theta}}^{(k)} - \left(\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}^{(k)})\right)^{-1} \dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}^{(k)})$$

- **Method of Scoring** For $k = 0, 1, \ldots$, define

$$\text{(i)} \quad \widehat{\underset{\sim}{\theta}}^{(k+1)} = \widehat{\underset{\sim}{\theta}}^{(k)} + \left(\mathcal{I}(\widehat{\underset{\sim}{\theta}}^{(k)})\right)^{-1} \left(\frac{1}{n}\,\dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}^{(k)})\right)$$

$$\text{(ii)} \quad \widehat{\underset{\sim}{\theta}}^{(k+1)} = \widehat{\underset{\sim}{\theta}}^{(k)} + \left(\widehat{I}_n(\widehat{\underset{\sim}{\theta}}^{(k)})\right)^{-1} \left(\frac{1}{n}\,\dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}^{(k)})\right)$$

  if $\mathcal{I}(.)$ is not known precisely. Note that

$$-\frac{1}{n}\ddot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}^{(k)}) \xrightarrow{p} \mathcal{I}(\underset{\sim}{\theta})$$

The following theorem proves that only **one** iterative step is required to match the asymptotic efficiency of solutions to the likelihood equations, which have been shown to have asymptotic variance equal to the Cramér-Rao information bound. The method of proof is as follows

1 Find a consistent but possibly inefficient estimator $\widetilde{\underset{\sim}{\theta}}_n$

2 Form the one-step Newton or Scoring Estimator using the formulae

$$\widehat{\underset{\sim}{\theta}}^{\dagger} = \widetilde{\underset{\sim}{\theta}}_n - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n) \qquad \widehat{\underset{\sim}{\theta}}^{\star} = \widetilde{\underset{\sim}{\theta}}_n + \left(\mathcal{I}(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \frac{1}{n} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)$$

3 Show that these estimators have the same asymptotic properties as solutions to the likelihood equations. That is, under the regularity conditions, if $\widehat{\underset{\sim}{\theta}}_n$ satisfies

$$\dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n) = \underset{\sim}{0} \tag{LE}$$

then,

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})$$

and the following theorem shows that $\widehat{\underset{\sim}{\theta}}^{\dagger}$ and $\widehat{\underset{\sim}{\theta}}^{\star}$ also have these properties.

Theorem The Efficiency of One-Step Estimators
Let $\widetilde{\underset{\sim}{\theta}}_n$, $n = 1, 2, \ldots$, be a consistent sequence of estimators of $\underset{\sim}{\theta}_0 \in \Theta$ such that

$$\sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \Sigma(\underset{\sim}{\theta}_0))$$

with $\Sigma(\underset{\sim}{\theta}_0)$ finite. Then, under the conditions A0-A4, the two estimators

$$\widehat{\underset{\sim}{\theta}}^{\dagger} = \widetilde{\underset{\sim}{\theta}}_n - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n) \tag{1}$$

$$\widehat{\underset{\sim}{\theta}}^{\star} = \widetilde{\underset{\sim}{\theta}}_n + \left(\mathcal{I}(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \frac{1}{n} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n) \tag{2}$$

are **asymptotically equivalent** to the MLE,

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{0} \qquad \sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\star} - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{0}$$

and

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}) \qquad \sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\star} - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})$$

Proof Suppose that $\widehat{\underset{\sim}{\theta}}_n$ is a consistent sequence of estimators that satisfy (LE), with

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})$$

Note: At no stage in the estimation will we actually have to find the numerical value of $\widehat{\underset{\sim}{\theta}}_n$.

A first-order Taylor expansion of $\dot{\underset{\sim}{l}}_n$ about $\widehat{\underset{\sim}{\theta}}_n$ yields the following equation:

$$\dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n) = \dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n) + \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) = \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n). \tag{3}$$

as $\widehat{\underset{\sim}{\theta}}_n$ is a solution to equation (LE), for some $\underset{\sim}{\theta}_n^{\ddagger}$ where

$$\|\underset{\sim}{\theta}_n^{\ddagger} - \widehat{\underset{\sim}{\theta}}_n\| \leq \|\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n\|$$

In equation (3), the left hand side is a $d \times 1$ vector, and $\ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})$ is a $d \times d$ matrix. By definition,

$$(\widehat{\underset{\sim}{\theta}}^{\dagger} - \widehat{\underset{\sim}{\theta}}_n) = (\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)$$

so that, by equation (3),

$$\begin{aligned}\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \widehat{\underset{\sim}{\theta}}_n) &= \sqrt{n}\left[(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \dot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right] = \sqrt{n}\left[(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n)\right] \\ &= \left[\mathbf{1}_d - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})\right] \sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) \qquad (4)\end{aligned}$$

where $\mathbf{1}_d$ is the $d \times d$ identity matrix. Recall that both $\widehat{\underset{\sim}{\theta}}_n$ and $\widetilde{\underset{\sim}{\theta}}_n$ are consistent by assumption

$$\widehat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}_0 \qquad \widetilde{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}_0 \qquad \therefore \qquad \widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{0}.$$

Furthermore, by the Weak Law of Large Numbers, and the continuous mapping theorem

$$\frac{1}{n}\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n) \xrightarrow{p} -\mathcal{I}(\underset{\sim}{\theta}_0)$$

and

$$\widehat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}_0 \qquad \text{and} \qquad \widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{0} \qquad \therefore \qquad \underset{\sim}{\theta}_n^{\ddagger} \xrightarrow{p} \underset{\sim}{\theta}_0$$

so therefore

$$\frac{1}{n}\ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger}) \xrightarrow{p} -\mathcal{I}(\underset{\sim}{\theta}_0).$$

Therefore, in equation (4)

$$\left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger}) \xrightarrow{p} \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}\mathcal{I}(\underset{\sim}{\theta}_0) = \mathbf{1}_d$$

and so

$$\mathbf{1}_d - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger}) \xrightarrow{p} \mathbf{1}_d - \mathbf{1}_d = \mathbf{0}_d \qquad (5)$$

Also in equation (4),

$$\sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) = \sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) - \sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0)$$

and, by assumption

$$\left.\begin{array}{rcl}\sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) & \xrightarrow{d} & N(0, \Sigma(\underset{\sim}{\theta}_0)) \\ \sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) & \xrightarrow{d} & N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})\end{array}\right\} \qquad \Longrightarrow \qquad \sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{d} \underset{\sim}{Z}_0 \sim N(0, \Sigma(\underset{\sim}{\theta}_0) + \mathcal{I}(\underset{\sim}{\theta}_0)^{-1})$$

Hence, from equations (4) and (5), and Slutsky's Theorem,

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \widehat{\underset{\sim}{\theta}}_n) = \left[\mathbf{1}_d - \left(\ddot{\underset{\sim}{l}}_n(\widetilde{\underset{\sim}{\theta}}_n)\right)^{-1} \ddot{\underset{\sim}{l}}_n(\underset{\sim}{\theta}_n^{\ddagger})\right] \sqrt{n}(\widetilde{\underset{\sim}{\theta}}_n - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \mathbf{0}_d \times \underset{\sim}{Z}_0 = \underset{\sim}{0}$$

Hence the two estimators are asymptotically equivalent, that is

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \widehat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{0}.$$

But the asymptotic distribution of $\sqrt{n}(\widehat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0)$ is known, and therefore

$$\sqrt{n}(\widehat{\underset{\sim}{\theta}}^{\dagger} - \underset{\sim}{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}(\underset{\sim}{\theta}_0)^{-1}).$$

Example: Cauchy(θ)

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

The MLE $\widehat{\theta}_n$ can only be computed numerically. An alternative estimator is the sample median $\widetilde{\theta}_n = \widetilde{x}$, say, which is consistent for θ. Then the Newton one-step estimator is

$$\underset{\sim}{\widehat{\theta}}^{\dagger} = \underset{\sim}{\widetilde{\theta}}_n - \frac{\dot{l}_n(\underset{\sim}{\widetilde{\theta}}_n)}{\ddot{l}_n(\underset{\sim}{\widetilde{\theta}}_n)} = \widetilde{x} - \frac{\displaystyle\sum_{i=1}^{n}\frac{(x_i-\widetilde{x})}{1+(x_i-\widetilde{x})^2}}{\displaystyle\sum_{i=1}^{n}\frac{(x_i-\widetilde{x})^2-1}{(1+(x_i-\widetilde{x})^2)^2}}.$$

The simulation below shows the variances estimated as the variance of T_n

$$T_n = \sqrt{n}(\theta_n - \theta)$$

for $\widehat{\theta}_n$, $\widetilde{\theta}_n$ and $\underset{\sim}{\widehat{\theta}}^{\dagger}$, for 50000 replicate data sets of size $n = 20, 40, 100$ and 10000.

n	Measure	$\widehat{\theta}_n$	$\widetilde{\theta}_n$	$\underset{\sim}{\widehat{\theta}}^{\dagger}$
20	Variance	2.313	2.765	6.191
	Efficiency	1.000	0.837	0.374
40	Variance	2.164	2.647	2.169
	Efficiency	1.000	0.818	0.998
100	Variance	2.035	2.503	2.036
	Efficiency	1.000	0.813	1.000
10000	Variance	2.010	2.480	2.010
	Efficiency	1.000	0.811	1.000

In this case

$$\begin{aligned} \mathcal{I}(\theta) &= \int_{-\infty}^{\infty}(S(x,\theta))^2 f_{X|\theta}(x|\theta)\,dx \\ &= \int_{-\infty}^{\infty}\frac{4(x-\theta)^2}{\pi(1+(x-\theta)^2)^3}\,dx \\ &= \int_{-\infty}^{\infty}\frac{4x^2}{\pi(1+x^2)^3}\,dx = \frac{1}{2} \end{aligned}$$

so the asymptotic variance of the efficient estimator is $\mathcal{I}(\theta)^{-1} = 2$.

Quadratic Approximation
A second-order Taylor *approximation* to the log-likelihood around the MLE, $\widehat{\underset{\sim}{\theta}}_n$, yields a quadratic approximation to the log-likelihood:

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

as $\dot{\underset{\sim}{l}}_n(\widehat{\underset{\sim}{\theta}}_n) = \underset{\sim}{0}$.
In the one-dimensional case, we have

$$l_n(\theta) \doteq l_n(\widehat{\theta}) + \frac{\ddot{l}_n(\widehat{\theta})}{2}(\theta - \widehat{\theta})^2$$

Example: Poisson(λ)
Let $s_n = \sum_{i=1}^{n} x_i$. Then

$$\begin{aligned} l_n(\lambda) &= \text{constant} + s_n \log \lambda - n\lambda \\ \dot{l}_n(\lambda) &= s_n/\lambda - n \\ \ddot{l}_n(\lambda) &= -s_n/\lambda^2 \end{aligned}$$

and as the MLE is $\widehat{\lambda}_n = \bar{x}$, we have the likelihood approximation

$$l_n(\lambda) = l_n(\widehat{\lambda}_n) - \frac{1}{2}\frac{s_n}{\widehat{\lambda}_n^2}(\lambda - \widehat{\lambda}_n)^2 = l_n(\bar{x}) - \frac{n(\lambda - \bar{x})^2}{2\bar{x}}$$

Example: Normal$(0, \sigma^2) \equiv$ Normal$(0, \theta)$.
Let $q_n = \sum_{i=1}^{n} x_i^2$, we have

$$\begin{aligned} l_n(\theta) &= \text{constant} - \frac{n}{2}\log\theta - \frac{q_n}{2\theta} \\ \dot{l}_n\theta) &= -\frac{n}{2\theta} + \frac{q_n}{2\theta^2} \\ \ddot{l}_n(\theta) &= \frac{n}{2\theta^2} - \frac{q_n}{\theta^3} \end{aligned}$$

The MLE is $\widehat{\theta}_n = q_n/n$, and thus

$$l_n(\widehat{\theta}_n) = \frac{n}{2\widehat{\theta}_n^2} - \frac{q_n}{\widehat{\theta}_n^3} = -\frac{n^3}{2q_n^2}$$

so we have the likelihood approximation

$$l_n(\theta) = \dot{l}_n(\widehat{\theta}_n) - \frac{1}{2}\frac{n^3}{2q_n^2}(\theta - \widehat{\theta}_n)^2 = l_n(q_n/n) - \frac{n^3(\theta - q_n/n)^2}{4q_n^2}.$$

5.2 Hypothesis Testing

Asymptotic behaviour of the Likelihood Statistic

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

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

is based on the statistic

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

say, where $\Theta \equiv \Theta_0 \cup \Theta_1$; H_0 is **rejected** if $\lambda_{\underset{\sim}{X}}(\underset{\sim}{x})$ is **small enough**, that is, $\lambda_{\underset{\sim}{X}}(\underset{\sim}{x}) \leq k$ for some k to be defined. **Note that Θ is not necessarily the entire parameter space, just the union of Θ_0 and Θ_1.**

Theorem Consider using the likelihood ratio statistic for testing

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

Then under assumptions A0-A4, the maximum likelihood estimator for θ under H_1, $\widehat{\theta}_n$, is consistent and an asymptotically normally distributed, and as $n \longrightarrow \infty$ under H_0,

$$-2 \log \lambda_{\underset{\sim}{X}}(\underset{\sim}{X}) = -2(l_n(\theta_0) - l_n(\widehat{\theta}_n)) \xrightarrow{d} Q \sim \chi^2_1$$

Proof A second-order Taylor expansion to the log-likelihood around the MLE, $\widehat{\theta}_n$

$$l_n(\theta) = l_n(\widehat{\theta}_n) + (\theta - \widehat{\theta}_n)\dot{l}_n(\widehat{\theta}_n) + \frac{1}{2}(\theta - \widehat{\theta}_n)^2 \ddot{l}_n(\widehat{\theta}_n) + \frac{1}{6}(\theta - \widehat{\theta}_n)^3 \dddot{l}_n(\theta^\star_n)$$

where $\theta^\star_n$ lies between θ and $\widehat{\theta}_n$. Now as $\widehat{\theta}_n$ is the maximum likelihood estimate

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

and therefore on rearrangement, evaluating at $\theta = \theta_0$,

$$-2(l_n(\theta_0) - l_n(\widehat{\theta}_n)) = -(\theta_0 - \widehat{\theta}_n)^2 \ddot{l}_n(\widehat{\theta}_n) - \frac{1}{3}(\theta_0 - \widehat{\theta}_n)^3 \dddot{l}_n(\theta^\star_n).$$

In this expression, consider the asymptotic behaviour in terms of the random variables. We have

$$-(\theta_0 - \widehat{\theta}_n)^2 \ddot{l}_n(\widehat{\theta}_n) = (\sqrt{n}(\theta_0 - \widehat{\theta}_n))^2 \left(-\frac{1}{n} \ddot{l}_n(\widehat{\theta}_n) \right)$$

where

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

and

$$-\frac{1}{n} \ddot{l}_n(\widehat{\theta}_n) \xrightarrow{p} \mathcal{I}(\theta_0)$$

so that, by Slutsky's Theorem

$$-(\theta_0 - \widehat{\theta}_n)^2 \ddot{l}_n(\widehat{\theta}_n) \xrightarrow{d} \chi^2_1$$

Finally,

$$(\theta_0 - \widehat{\theta}_n)^3 \dddot{l}_n(\theta^\star_n) = (\theta_0 - \widehat{\theta}_n) \left((\theta_0 - \widehat{\theta}_n)^2 \dddot{l}_n(\theta^\star_n) \right) \xrightarrow{p} 0$$

so

$$-2(l_n(\theta_0) - l_n(\widehat{\theta}_n)) \overset{d}{=} Q + \mathrm{o}_P(1) \xrightarrow{d} Q,$$

where $Q \sim \chi^2_1$. ■