Asymptotic Properties of M.L.E.'s

We have seen that in finite samples, MLE's. may have "good" or "bad" properties. Their main justification is that, subject to certain "regularity conditions" on the underlying data density, they have good <u>asymptotic</u> properties. They are:

* Weakly (& often strongly) consistent

* Asymptotically Normal

* Asymptotically Efficient.

We'll prove these results.

First, let's consider, more formally, the whole notion of estimator efficiency (& hence asymptotic efficiency).

<u>Theorem</u> : (Cramér – Rao)

Let $\hat{\theta}$ be an estimator of the scalar, θ

Suppose $E(\hat{\theta})$ is finite, & differentiable w.r.t. θ. Then:

$$\text{Var.}(\hat{\theta}) \geq -\left[\frac{\partial E(\hat{\theta})}{\partial \theta}\right]^2 / E[H(\theta)]$$

where $H(\cdot)$ is the Hessian for the log-likelihood fctn.

An estimator that attains this lower bound is efficient, but the bound may be unattainable. (**Note – likelihood fctn**

In the case where θ is a <u>vector</u> –

$[V(\hat{\theta}) - B\, I(\theta)^{-1} B']$ is p.s.d., where

$$\begin{cases} B = (\partial E(\hat{\theta}) / \partial \theta) \\ I(\theta) = -E[H(\theta)] = -E\left[\dfrac{\partial^2 \log L(\theta)}{\partial \theta\, \partial \theta'}\right] \end{cases}$$

Fisher's "Information Matrix"

N.B.: If $\hat{\theta}$ is unbiased, this last result becomes: $[V(\hat{\theta}) - I(\theta)^{-1}]$ is p.s.d.

Theorem: Any unbiased estimator that attains the C.R.L.B. is an M.L.E.

(Only limited practical usefulness.)

Definition: The Asymptotic Information Matrix is: $IA(\theta) = \lim_{n \to \infty} [\frac{1}{n} I(\theta)]$

So, any asymptotically unbiased estimator which has an asymptotic covariance matrix equal to $IA(\cdot)^{-1}$ must be Asymptotically Efficient.

As we'll see – MLE's are asymptotically efficient, & in fact if $\tilde{\theta}$ is an MLE,

$$\sqrt{n}\,(\tilde{\theta} - \theta) \xrightarrow{d} N[0, (IA(\theta))^{-1}].$$

MLE's are "Best Asymptotically Normal" (BAN)

Our first major task is to prove the Cramér-Rao Theorem (scalar case).

Theorem: (Cramér-Rao)

If $\hat{\theta}$ is an estimator of (scalar) θ, where $E(\hat{\theta})$ exists (finite) & is differentiable w.r.t. θ, then:

$$\text{var.}(\hat{\theta}) \geq -\left[\frac{\partial E(\hat{\theta})}{\partial \theta}\right]^2 \Big/ E[H(\theta)]$$

where $H(\theta) = \left(\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right)$.

Proof: First, let's establish a preliminary result. Note that

$$\int L(\theta|y)\,dy = \int p(y|\theta)\,dy = 1$$

so: $$\frac{\partial}{\partial \theta}\int L(\theta|y) = \frac{\partial}{\partial \theta}(1) = 0$$

or, $$\int \frac{\partial \log L(\theta|y)}{\partial \theta} \cdot L(\theta|y)\,dy = 0$$

or, $$\int \frac{\partial \log L(\theta|y)}{\partial \theta}\, p(y|\theta)\,dy = 0$$

or, $$E\left[\frac{\partial \log L(\theta|y)}{\partial \theta}\right] = 0. \qquad \#.$$

Now: $E(\hat{\theta})$ is finite & differentiable, so

$$\frac{\partial}{\partial\theta}E(\hat{\theta}) = \frac{\partial}{\partial\theta}\int \hat{\theta}\, p(y|\theta)\,dy$$

$$= \int \hat{\theta}\, \frac{\partial L(\theta|y)}{\partial\theta}\,dy$$

$$= \int \hat{\theta}\, \frac{\partial \log L(\theta|y)}{\partial\theta}\, L(\theta|y)\,dy$$

$$= \int \hat{\theta}\left[\frac{\partial \log L(\theta|y)}{\partial\theta} - 0\right] L(\theta|y)\,dy$$

$$= E\left[\hat{\theta}\left[\frac{\partial \log L(\theta|y)}{\partial\theta} - 0\right]\right]$$

$$= E\left\{\left[\hat{\theta} - E(\hat{\theta})\right]\left[\frac{\partial \log L(\theta)}{\partial\theta} - 0\right]\right\}$$

$$+ E\left\{E(\hat{\theta})\left[\frac{\partial \log L(\theta)}{\partial\theta} - 0\right]\right\}$$

$$= \text{cov.}\left\{\hat{\theta}, \frac{\partial \log L(\theta)}{\partial\theta}\right\} + E(\hat{\theta}).\underbrace{E\left[\frac{\partial \log L(\theta)}{\partial\theta}\right]}_{=0}.$$

$$= \text{cov.}\left\{\hat{\theta}, \frac{\partial \log L(\theta)}{\partial\theta}\right\}.$$

So, by the <u>Cauchy - Schwarz Inequality</u>,

$$[\text{cov.}(x,y)]^2 \le [\text{var.}(x)][\text{var.}(y)]$$

& so

$$\left[\frac{\partial E(\hat{\theta})}{\partial\theta}\right]^2 \le \text{var.}(\hat{\theta}).\,\text{var.}\left(\frac{\partial \log L(\theta)}{\partial\theta}\right)$$

or,

$$\text{Var.}(\hat{\theta}) \geq \left[\frac{\partial E(\hat{\theta})}{\partial\theta}\right]^2 \Big/ \text{var.}\left(\frac{\partial \log L(\theta)}{\partial\theta}\right)$$

We now need to show that

$$\text{var.}\left[\frac{\partial \log L(\theta)}{\partial\theta}\right] = -E\left[\frac{\partial^2 \log L(\theta)}{\partial\theta^2}\right].$$

To see this — we showed already that

$$E\left[\frac{\partial \log L(\theta)}{\partial\theta}\right] = \int \frac{\partial \log L(\theta)}{\partial\theta}\, L(\theta|y)\,dy = 0$$

Differentiating again:

$$\int\left[\frac{\partial^2 \log L(\theta)}{\partial\theta^2} L(\theta) + \frac{\partial \log L(\theta)}{\partial\theta}.\frac{\partial L(\theta)}{\theta}\right]dy = 0$$

So:

$$-\int \frac{\partial^2 \log L(\theta)}{\partial\theta^2} L(\theta)\,dy = \int \frac{\partial \log L(\theta)}{\partial\theta}.\frac{\partial L(\theta)}{\partial\theta}\,dy$$

$$= \int \frac{\partial \log L(\theta)}{\partial\theta}\,\frac{\partial \log L(\theta)}{\partial\theta}.\frac{\partial L(\theta)}{\partial \log L(\theta)}\,dy$$

$$= \int\left[\frac{\partial \log L(\theta)}{\partial\theta}\right]^2 L(\theta)\,dy$$

$$= \int\left[\frac{\partial \log L(\theta)}{\partial\theta} - 0\right]^2 p(y|\theta)\,dy$$

$$= \text{var.}\left(\frac{\partial \log L(\theta)}{\partial\theta}\right).$$

i.e.

$$\text{var.}\left(\frac{\partial \log L(\theta)}{\partial\theta}\right) = -E\left[\frac{\partial^2 \log L(\theta)}{\partial\theta^2}\right]$$

& So:

$$\text{var.}(\hat{\theta}) \geq -\left[\frac{\partial E(\hat{\theta})}{\partial\theta}\right]^2 \Big/ E[H(\theta)].$$

$(\geq -1/E[H(\theta)]$, if unbiased

<u>Example:</u>

$$y = X\beta + \varepsilon \quad ; \varepsilon \sim N[0, \sigma^2 I]$$

$$L(\beta, \sigma^2 | y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}$$

$$\log L = -(n/2)\log(\sigma^2) - n/2 \log(2\pi) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

(i) $(\partial \log L / \partial \beta) = -\frac{1}{2\sigma^2}(2X'X\beta - 2X'y)$

(ii) $(\partial \log L / \partial \sigma^2) = -n/2\sigma^2 + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta)$

$$\Rightarrow \tilde{\beta} = (X'X)^{-1}X'y \quad ; \tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta}).$$

(iii) $(\partial^2 \log L / \partial\beta\partial\beta') = -(X'X)/\sigma^2$

(iv) $(\partial^2 \log L / \partial\sigma^2\partial\sigma^2) = n/2\sigma^4 - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta)$

(v) $(\partial^2 \log L / \partial\beta\partial\sigma^2) = 1/\sigma^4 (X'X\beta - X'y)$

From (iii): $-E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right] = (X'X)/\sigma^2$

From (iv): $-E\left[\frac{\partial^2 \log L}{\partial\sigma^2\partial\sigma^2}\right] = -n/2\sigma^4 + \frac{1}{\sigma^6}E(\varepsilon'\varepsilon)$

$$= -n/2\sigma^4 + \frac{1}{\sigma^6}E\left(\sum_i \varepsilon_i^2\right)$$

$$= -n/2\sigma^4 + \frac{n\sigma^2}{\sigma^6}$$

$$= n/2\sigma^4.$$

From (v):

$$-E\left[\frac{\partial^2 \log L}{\partial\beta\partial\sigma^2}\right] = -1/\sigma^4 (X'X\beta - X'E(y)) = 0.$$

So, Fisher's Information Matrix is:

$$I(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \frac{X'X}{\sigma^2} & \underset{\sim}{0} \\ \underset{\sim}{0} & n/2\sigma^4 \end{bmatrix}$$

$$I^{-1}(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \sigma^2(X'X)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & 2\sigma^4/n \end{bmatrix}$$

$$V(\tilde{\beta}) = \sigma^2(X'X)^{-1} = \text{CRLB for } \underset{\sim}{\beta}.$$

So, with Normal errors, $\tilde{\beta}$ is Best Unbiased.

Consider the asymptotic case –

$$\frac{1}{n}I(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \frac{1}{n}(X'X)/\sigma^2 & \underset{\sim}{0} \\ \underset{\sim}{0} & 1/2\sigma^4 \end{bmatrix}$$

So, $IA(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} Q/\sigma^2 & \underset{\sim}{0} \\ \underset{\sim}{0} & 1/2\sigma^4 \end{bmatrix}$

where $Q = \lim_{n\to\infty}\left[\frac{1}{n}X'X\right]$

We mentioned that

$$\sqrt{n}(\tilde{\theta} \rightarrow \theta) \xrightarrow{d} N[0, IA^{-1}]$$

So, Fisher's Information Matrix is:

$$I(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \frac{X'X}{\sigma^2} & \underset{\sim}{0} \\ \underset{\sim}{0} & n/2\sigma^4 \end{bmatrix}$$

$$I^{-1}(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \sigma^2 (X'X)^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & 2\sigma^4/n \end{bmatrix}$$

$V(\tilde{\beta}) = \sigma^2 (X'X)^{-1} = \text{CRLB for } \underset{\sim}{\beta}.$

So, with Normal errors, $\tilde{\beta}$ is Best Unbiased.

Consider the asymptotic case –

$$\frac{1}{n} I(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} \frac{1}{n}(X'X)/\sigma^2 & \underset{\sim}{0} \\ \underset{\sim}{0} & 1/2\sigma^4 \end{bmatrix}$$

So, $IA(\underset{\sim}{\beta}, \sigma^2) = \begin{bmatrix} Q/\sigma^2 & \underset{\sim}{0} \\ \underset{\sim}{0} & 1/2\sigma^4 \end{bmatrix}$

where $Q = \underset{n\to\infty}{\text{Limit}} \left[\frac{1}{n} X'X \right]$

We mentioned that

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N[0, IA^{-1}]$$

So: $\begin{cases} \sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} N[0, \sigma^2 Q^{-1}] \\ \sqrt{n}(\tilde{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4) \end{cases}$

because they are MLE's.

N.B.:

Asy. var. $[\sqrt{n}(\tilde{\theta} - \theta)] = IA^{-1}(\theta)$

Asy. var. $(\tilde{\theta}) = \frac{1}{n} IA^{-1}(\theta)$

Est. asy. var. $(\tilde{\theta}) = \frac{1}{n} IA^{-1}(\tilde{\theta})$

So,

$$\frac{\tilde{\theta}_i - \theta_i}{\text{a.s.e.}(\tilde{\theta}_i)} \xrightarrow{d} N(0, 1).$$

("Asymptotic t-statistics".)

To get Est. Asy. var. $(\tilde{\beta})$, use $\tilde{\sigma}^2 (X'X)^{-1}$.

because: $\text{plim}(\tilde{\sigma}^2) = \sigma^2$

$\text{plim}(n(X'X)^{-1}) = \text{plim}\left[\left(\frac{1}{n} X'X\right)^{-1} \right] = Q^{-1}$

So $\text{plim}(n\tilde{\sigma}^2 (X'X)^{-1}) = \sigma^2 Q^{-1}$.

Further Examples:

Binomial Dist'n –

$$L(p \mid x, n) = p(x \mid p, n) = {}^nC_x \, p^x (1-p)^{n-x}$$

$$= k \, p^x (1-p)^{n-x}$$

So,

$$\log L = k' + x \log p + (n-x) \log(1-p)$$

$$(\partial \log L / \partial p) = x/p - \left(\frac{n-x}{1-p} \right) = 0$$

$$\Rightarrow \tilde{p} = x/n$$

$$(\partial^2 \log L / \partial p^2) = - x/p^2 - \frac{(n-x)}{(1-p)^2} \qquad (< 0)$$

$$= H$$

Now,

$$I(p) = - E\{H\}$$

$$= - \left[- \frac{np}{p^2} - \frac{(n-np)}{(1-p)^2} \right]$$

$$= \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)}$$

$$IA(p) = \lim_{n \to \infty} \left[\frac{1}{n} I(p) \right] = \frac{1}{p(1-p)}$$

So,

$$\sqrt{n}\,(\tilde{p} - p) \xrightarrow{d} N[0, p(1-p)]$$

Poisson Dist'n –

$$p(x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \quad ; \quad x_i = 0, 1, 2, \ldots \quad \lambda > 0.$$

N.B.: $$E(x_i) = \sum_{i=0}^{\infty} x_i \, p(x_i)$$
$$= \sum_{i=0}^{\infty} x_i \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$
$$= \sum_{i=1}^{\infty} \frac{\lambda^{x_i} e^{-\lambda}}{(x_i-1)!}$$
$$= \lambda \sum_{i=1}^{\infty} \frac{\lambda^{x_i-1} e^{-\lambda}}{(x_i-1)!}$$
$$= \lambda \sum_{i=0}^{\infty} \frac{\lambda^{x_i'} e^{-\lambda}}{x_i'!} \quad ; \quad x_i' = x_i - 1$$
$$= \lambda . 1 = \lambda.$$

Now, $$L(\lambda | \underset{\sim}{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{(x_i)!}$$
$$= \frac{\lambda^{\Sigma x_i} e^{-n\lambda}}{\prod_i x_i!}$$

$$\log L(\lambda | \underset{\sim}{x}) = \sum_i x_i \log \lambda - n\lambda - k$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{\Sigma x_i}{\lambda} - n = 0 \; ; \; \Rightarrow \underline{\tilde{\lambda} = \frac{1}{n} \sum_i x_i}$$

& $$\frac{\partial^2 \log L}{\partial \lambda^2} = - \sum_i x_i / \lambda^2 \quad (<0)$$
$$= H(\lambda)$$

$$I(\lambda) = -E[H(\lambda)] = \frac{1}{\lambda^2} \sum_i E(x_i)$$
$$= \frac{1}{\lambda^2} \cdot n\lambda = n/\lambda.$$

So, $$IA(\lambda) = \lim_{n \to \infty} \left[\frac{1}{n} \cdot \frac{n}{\lambda} \right] = 1/\lambda$$

& so $$\sqrt{n}(\tilde{\lambda} - \lambda) \xrightarrow{d} N[0, \lambda].$$

Pareto Distribution:

$$p(x_i) = c \, x_i^{-(c+1)} \quad ; \; c > 0 \; ; \; 1 < x_i < \infty$$

Show that $$\left\{ \begin{array}{l} \tilde{c} = (n / \sum_i \log x_i) \\ I = n/c^2 \\ IA = 1/c^2 \end{array} \right.$$

So, $$\sqrt{n}(\tilde{c} - c) \xrightarrow{d} N(0, c^2)$$

& $$\text{a.s.e.}(\tilde{c}) = \tilde{c}/\sqrt{n}.$$

We have noted already that MLE's are weakly consistent & BAN. Let's prove these 2 results. We'll look at the i.i.d. scalar case, where the density satisfies the following Regularity Conditions (on $f_i = f(y_i|\theta)$)

1. $\dfrac{\partial^{(n)} \log f_i}{\partial \theta^{(n)}}$ exist for $n = 1, 2, 3.$

2. $\displaystyle\int \frac{\partial \log f_i}{\partial \theta} \cdot dy = \frac{\partial}{\partial \theta} \int \log f_i \, dy.$

   (can differentiate under integration O.K.)

3. $\left.\dfrac{\partial \log f_i}{\partial \theta}\right|_{\theta=\theta_0}$ has positive & finite variance.

4. $\dfrac{\partial^3 \log f_i}{\partial \theta^3}$ is bounded.

(These conditions are sometimes expressed differently.)

<u>Theorem</u>: Suppose we have i.i.d. sampling. If the likelihood equations have a unique root, the MLE is weakly consistent.

<u>Proof</u>: (Scalar case)

$$\log L(\theta|\underline{y}) = \sum_i \log f(y_i|\theta)$$

We obtain MLE by solving $\dfrac{\partial \log L}{\partial \theta} = 0$, for θ.

i.e. solve $\displaystyle\sum_i \partial \log f(y_i|\theta)/\partial\theta = 0.$

Now,

$$\frac{\partial \log f(y_i|\theta)}{\partial \theta} = \left.\left(\frac{\partial \log f_i}{\partial \theta}\right)\right|_{\theta_0} + (\theta-\theta_0)\left.\left(\frac{\partial^2 \log f_i}{\partial \theta^2}\right)\right|_{\theta_0} + \frac{1}{2}(\theta-\theta_0)^2\left.\left(\frac{\partial^3 \log f_i}{\partial \theta^3}\right)\right|_{\theta_0} + \cdots$$

So,

$$\frac{1}{n}\left(\frac{\partial \log L}{\partial \theta}\right) \approx B_0(y) + (\theta-\theta_0)B_1(y) + \frac{1}{2}(\theta-\theta_0^2)B_2(y)$$

(Note that this expansion is valid only if θ_0 is an interior point of parameter space. So, if θ_0 is a boundary point, the proof fails, & we don't get consistency for the MLE.)

Now, $B_0(y) = \frac{1}{n} \sum_i \left(\frac{\partial \log f_i}{\partial \theta} \right) \Big|_{\theta_0}$

$B_1(y) = \frac{1}{n} \sum_i \left(\frac{\partial^2 \log f_i}{\partial \theta^2} \right) \Big|_{\theta_0}$

$B_2(y) = \frac{1}{n} \sum_i \left(\frac{\partial^3 \log f_i}{\partial \theta^3} \right) \Big|_{\theta_0}$

(& B_2 is bounded)

These sample averages converge in probability to their population means, by Khintchine's Theorem.

Now, in proving the Cramér-Rao Theorem, we showed that $E\left[\frac{\partial \log L}{\partial \theta} \right] = 0$. (& set $n=1$)

So, $B_0(y) \xrightarrow{p} 0$.

Also, $-E\left[\frac{\partial^2 \log L}{\partial \theta^2} \right] = \text{var.}\left(\frac{\partial \log L}{\partial \theta} \right) = E\left(\frac{\partial \log L}{\partial \theta} \right)^2$.

So, $B_1(y) \xrightarrow{p} E\left[\left(\frac{\partial \log f_i}{\partial \theta} \right)^2 \right] = -E\left[\frac{\partial^2 \log f}{\partial \theta^2} \right] = -k^2$.

Because $B_2(y)$ is bounded,

$\text{plim}\,(B_2(y)) = \xi M$; $0 \le \xi < 1$

Consider the RHS of $*$:

let $\theta = \theta_0 \mp \delta$, for small $\delta > 0$.

$$
\begin{array}{ccccccc}
\text{RHS} = & B_0(y) & \pm & \delta B_1(y) & + & \frac{1}{2}\delta^2 B_2(y) \\
& \downarrow p & & \downarrow p & & \downarrow p \\
& 0 & & \mp k^2 \delta & & \frac{1}{2}\delta^2 \xi M
\end{array}
$$

By choosing δ small enough and n large enough, we can ensure the sum of the first & third terms is smaller in absolute value than the second term with probability $\to 1$.

So, the sign of $\frac{\partial \log L}{\partial \theta}$ is determined by the sign of the 2nd term — it is +ve, with arbitrarily high probability if $\theta = \theta_0 - \delta$, and -ve if $\theta = \theta_0 + \delta$.

Since $\frac{\partial^2 \log L}{\partial \theta^2}$ exists the likelihood equation, $(\partial \log L / \partial \theta) = 0$, has a root between the limits $\theta_0 \pm \delta$ with prob. $\to 1$ if δ small & n large. That is, this root is <u>consistent</u> for θ_0. #.

Some Issues:

1. Generally do **not** get consistency if:

   (a) θ_0 is a boundary point.

   (b) No. of parameters $\uparrow$ at rate that $n \uparrow$ (or greater).

2. Multiple roots for likelihood equation –

   In this case there may be several consistent roots. However, if both $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are both consistent roots, then can be shown to be asymptotically equivalent –

   $\sqrt{n}\,(\tilde{\theta}_1 - \tilde{\theta}_2) \xrightarrow{p} 0.$

   Also, under some additional conditions we can show that if $\tilde{\theta}$ corresponds to a global maximum of $\log L$ then it will be consistent, with probability one.

Theorem: Under the same regularity conditions,

$$\sqrt{n}\,(\tilde{\theta} - \theta_0) \xrightarrow{d} N\left[0,\ IA^{-1}(\theta)\right]$$

Proof: From $*$:

$$B_0(y) = (\tilde{\theta} - \theta_0)\left[-B_1(y) - \tfrac{1}{2}(\tilde{\theta} - \theta_0)B_2(y)\right]$$

or, $$\sqrt{n}\,(\tilde{\theta} - \theta_0) = \frac{\sqrt{n}\,B_0(y)}{-B_1(y) - \frac{1}{2}(\tilde{\theta} - \theta_0)B_2(y)}$$

or, $$k\sqrt{n}\,(\tilde{\theta} - \theta_0) = \left\{ \frac{\frac{1}{k\sqrt{n}} \sum_i \left(\frac{\partial \log f_i}{\partial \theta}\right)\Big|_{\theta_0}}{\frac{-B_1(y)}{k^2} - \frac{1}{2}\frac{(\tilde{\theta} - \theta_0)B_2(y)}{k^2}} \right\}$$

Now, $\text{plim}\,(-B_1/k^2) = 1.$

$B_2(y) \xrightarrow{p}$ finite value

& $\text{plim}\,(\tilde{\theta} - \theta_0) = 0.$

So, asy. dist'n. of $k\sqrt{n}\,(\tilde{\theta} - \theta_0)$ is same as that of $\frac{1}{k\sqrt{n}} \sum_i \left(\frac{\partial \log f_i}{\partial \theta}\right)\Big|_{\theta_0}.$

Sample average of iid. terms, $\left(\frac{\partial \log f_i}{\partial \theta}\right)$. Khintchine's Theorem, using $E\left[\frac{\partial \log f_i}{\partial \theta}\right] = 0$ & $\text{var.}\left(\partial \log f_i / \partial \theta\right) = k^2.$

So, by C.L.T. –

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

So,
$$\left[\frac{\frac{1}{n} \sum_i \left(\frac{\partial \log f_i}{\partial \theta} \right)\Big|_{\theta_0} - 0}{k/\sqrt{n}} \right] \xrightarrow{d} N(0,1)$$

or,
$$\frac{1}{k\sqrt{n}} \sum_i \left(\frac{\partial \log f_i}{\partial \theta} \right)\Big|_{\theta_0} \xrightarrow{d} N(0,1)$$

So,
$$\boxed{k\sqrt{n}\,(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0,1)}$$

Now,
$$k^2 = -E\left[\frac{\partial^2 \log f_i}{\partial \theta^2} \right]_{\theta_0}$$

$$nk^2 = -E\left[\frac{\partial^2 \log L}{\partial \theta^2} \right]_{\theta_0} = I(\theta)$$

& $\frac{1}{n} I(\theta) = k^2$, so

$$IA(\theta) = \lim_{n \to \infty} \left(\tfrac{1}{n} I(\theta) \right) = \lim_{n \to \infty} (k^2).$$

This implies that

$$\boxed{\sqrt{n}\,(\tilde{\theta} - \theta_0) \xrightarrow{d} N\left[0, IA^{-1}(\theta)\right]}$$