Chapter 8

Lecture 28

Example 7. X_i are iid uniformly over $(0, \theta)$ for $\theta \in \Theta = (0, \infty)$.

Homework 6

1. Show that

   a. With respect to Lebesgue measure on $\mathbb{R}^n$,

$$\ell(\theta \mid s_n) = \begin{cases} 1/\theta^n & \text{if } \theta \geq X_i \ \forall i \\ 0 & \text{otherwise} \end{cases}$$

   and $\hat{\theta} = \max\{X_1, \ldots, X_n\}$.

   b. Condition 2 in the Theorem above is satisfied, and hence $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta$ for all θ (which we check directly also); but the likelihood function is not continuous, and hence the information function is not defined.

   c. $E_\theta(\hat{\theta}_n) = \frac{n}{n+1}\theta$, and $\theta_n^* := \frac{n+1}{n}\hat{\theta}$ is unbiased.

   d. $n(\theta - \hat{\theta}_n)$ has the asymptotic distribution with density $\frac{1}{\theta}e^{-\frac{z}{\theta}}$ on $(0, \infty)$, and so $\hat{\theta}_n$ has a non-normal limiting distribution and $\hat{\theta}_n - \theta = O(1/n)$.

(In regular cases, $\hat{\theta}$ has a normal limiting distribution and $\hat{\theta}_n - \theta = O(1/\sqrt{n})$.)

Asymptotic distribution of $\hat{\theta}$ (θ real) in regular cases

$X = \{x\}$ (arbitrary), $\mathcal{C}$ is a σ-field on X, P_θ is a probability on $\mathcal{C}$ and $\theta \in \Theta$ for Θ an open interval in $\mathbb{R}^1$. $dP_\theta(x) = \ell(\theta \mid x)d\nu(x)$, with ν a fixed measure. Let $s_n = (X_1, \ldots, X_n) \in S^{(n)} = X \times \cdots \times X$, $\mathcal{A}^{(n)} = \mathcal{C} \times \cdots \times \mathcal{C}$ and $P_\theta^{(n)} = P_\theta \times \cdots \times P_\theta$ on $\mathcal{A}^{(n)}$. We assume that $\ell(\theta \mid x) > 0$, $L(\theta \mid x) = \log_e \ell(\theta \mid x)$ has at least two continuous derivatives, $E_\theta\big(L'(\theta \mid x)\big) = 0$ and

$$I_1(\theta) = E_\theta\big(L'(\theta \mid x)\big)^2 = -E_\theta\big(L''(\theta \mid x)\big) > 0.$$

We have $L(\theta \mid s_n) = \sum_{i=1}^n L(\theta \mid X_i)$, $L'(\theta \mid s_n) = \sum_{i=1}^n L'(\theta \mid X_i)$ and $L''(\theta \mid s_n) = \sum_{i=1}^n L''(\theta \mid X_i)$. For any given θ, we know that a good estimate of θ based on s_n will be approximately $a(\theta) + b(\theta)L'(\theta \mid s_n)$, and $L'(\theta \mid s_n) \approx N(0, *)$, so a good estimate of θ based on s_n will be approximately normally distributed when n is large. We have $\frac{L''(\theta|s_n)}{n} \to -I_1(\theta)$. Assume that:

Condition (*). Given any $\theta \in \Theta$, we may find an $\varepsilon = \varepsilon(\theta) > 0$ such that

$$\max_{|\delta-\theta|\leq\varepsilon} |L''(\delta \mid x)|$$

has a finite expectation under P_θ.

Assume also that $\hat{\theta}_n$ exists and is consistent. Then

$$0 = L'(\hat{\theta}_n \mid s_n) = L'(\theta \mid s_n) + (\hat{\theta}_n - \theta)L''(\theta_n^* \mid s_n),$$

where θ_n^* is between θ and $\hat{\theta}_n$. Since $\theta_n^* \to \theta$ in P_θ, we have

$$\left|\frac{L''(\theta_n^* \mid s_n)}{n} + I_1(\theta)\right| \to 0 \quad \text{in } P_\theta. \tag{**}$$

So

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{L'(\theta \mid s_n)}{\sqrt{n}} \cdot \frac{1}{I_1(\theta) + \xi_n},$$

where $\xi_n \to 0$ in P_θ. Since

$$\frac{L'(\theta \mid s_n)}{\sqrt{n}} \to N\big(0, I_1(\theta)\big) \quad \text{in distribution under } P_\theta,$$

we have:

1 (Fisher). $\sqrt{n}(\hat{\theta}_n - \theta) \to N\big(0, I_1(\theta)\big)$.

Note. This does *not* assert that $E_\theta(\hat{\theta}_n) = \theta + o(1)$ or that $\mathrm{Var}_\theta(\hat{\theta}_n) = \frac{1}{nI_1(\theta)} + o\big(\frac{1}{n}\big)$.

Proof of (**). Fix θ. Under (*), we have

$$h(r) := E_\theta\Big[\max_{|\delta-\theta|\leq r}\big|L''(\delta \mid x) - L''(\theta \mid x)\big|\Big] < +\infty$$

for sufficiently small $r > 0$. h is continuous in r and decreases to 0 as $r \to 0$.

For any $\eta > 0$, choose r such that $h(r) < \eta$. We have

$$\frac{1}{n}L''(\theta_n^* \mid s_n) = \frac{1}{n}L''(\theta \mid s_n) + \Delta_n,$$

where

$$|\Delta_n| = \frac{1}{n}\Big|\sum_{i=1}^n \big[L''(\theta_n^* \mid X_i) - L''(\theta \mid X_i)\big]\Big| \leq \frac{1}{n}\sum_{i=1}^n \big|L''(\theta_n^* \mid X_i) - L''(\theta \mid X_i)\big|.$$

Suppose that $|\hat{\theta}_n - \theta| < r$; then $|\theta_n^* - \theta| < r$ and hence $|\Delta_n| \le \frac{1}{n}\sum_{i=1}^n M(X_i)$, where $M(X) = \max_{|\delta-\theta|\le r}\left|L''(\delta \mid X) - L''(\theta \mid X)\right|$.

Since $E\left[M(X_i)\right] < \eta$, we have

$$\frac{1}{n}\sum_{i=1}^n M(X_i) \xrightarrow{\text{a.s.}} E_\theta\left[M(X)\right] < \eta.$$

Since η is arbitrary and $\hat{\theta}_n \to \theta$ in P_θ, we have that $|\Delta_n| \to 0$ in P_θ. □

Note. It was asserted by Fisher (and believed for a long time) that, if $t_n = t_n(s_n)$ is any estimate of θ such that

$$\sqrt{n}(t_n - \theta) \to N\big(0, v(\theta)\big) \quad \text{in distribution as } n \to \infty,$$

then $v(\theta) \ge 1/I_1(\theta)$. This is, however, not quite correct, as shown by the following counterexample (due to J. L. Hodges, 1951): Let X_i be iid $N(\theta, 1)$ and $\Theta = \mathbb{R}^1$. Let $\hat{\theta}_n = \overline{X_n}$. $\sqrt{n}(\hat{\theta}_n - \theta)$ is $N(0,1)$ and $I_1(\theta) = 1$. Let

$$t_n = \begin{cases} \overline{X_n} & \text{if } |\overline{X_n}| > n^{-1/4} \\ c\overline{X_n} & \text{if } |\overline{X_n}| \le n^{-1/4}; \end{cases}$$

then $\sqrt{n}(t_n - \theta) \to N\big(0, v(\theta)\big)$ for all θ, where

$$v(\theta) = \begin{cases} 1 & \text{if } \theta \ne 0 \\ c^2 & \text{if } \theta = 0, \end{cases}$$

and so $v(\theta) \ge 1$ breaks down at $\theta = 0$ (if we choose $-1 < c < 1$).

Lecture 29

Definition. We say that $\{z_n\}$ is $AN(\mu_n, \sigma_n^2)$ if

$$P\left(\frac{z_n - \mu_n}{\sigma_n} \le z\right) \to \Phi(z) \quad \text{for all } z.$$

Consider the condition

Condition (***). $\{t_n - \theta\}$ is $AN\big(0, v(\theta)/n\big)$ under θ (for each θ).

In Hodges's counterexample in the context of Example 1(a),

$$\sqrt{n}(t_n - \theta) = \varphi(\theta)\sqrt{n}(\overline{X_n} - \theta) + \xi_n(s, \theta),$$

where $\xi_n \to 0$ in P_θ-probability and

$$\varphi(\theta) = \begin{cases} 1 & \text{if } \theta \ne 0 \\ c & \text{if } \theta = 0, \end{cases}$$

so that t_n is $AN\big(\theta, v(\theta)/n\big)$ for $v(\theta) = \varphi^2(\theta)$. This provides an example of the following theorem:

2 (Le Cam/Bahadur). The set

$$\left\{\theta : v(\theta) < \frac{1}{I_1(\theta)}\right\}$$

is always of Lebesgue measure zero for any t_n satisfying (***).

Corollary. *If $\{t_n\}$ is regular in the sense that v is continuous in Θ and I_1 is also continuous, then $v(\theta) \geq 1/I_1(\theta)$ for all $\theta \in \Theta$.*

Note. This should not be confused with the C-R bound, since (***) does *not* imply that t_n is unbiased, nor that $v(\theta) \cong n \operatorname{Var}_\theta(t_n)$.

In the general case, (***) does imply that t_n is asymptotically median unbiased, i.e., that $P_\theta(t_n \leq \theta) \to \frac{1}{2}$ as $n \to \infty$ for each θ. Suppose this holds uniformly; then also it must be true that $v(\theta) \geq 1/I_1(\theta)$ for all θ. This follows from:

3. If θ is a point in Θ, $a > 0$ and $\delta_n(a) = \theta + \frac{a}{\sqrt{n}}$, and

$$\overline{\lim_{n\to\infty}}\, P_{\delta_n(a)}\big(t_n > \delta_n(a)\big) \geq \frac{1}{2},$$

then $v(\theta) \geq 1/I_1(\theta)$ (for the given θ).

Corollary. *Suppose that t_n is super-efficient ($v < 1/I_1$) at a point θ. Then, given any $a > 0$, we may find $\varepsilon_1 = \varepsilon_1(a) > 0$ and $\varepsilon_2 = \varepsilon_2(a) > 0$ such that*

$$P_{\theta+\frac{a}{\sqrt{n}}}\left(t_n > \theta + \frac{a}{\sqrt{n}}\right) < \frac{1}{2} - \varepsilon_1 \quad \textit{and} \quad P_{\theta-\frac{a}{\sqrt{n}}}\left(t_n < \theta - \frac{a}{\sqrt{n}}\right) < \frac{1}{2} - \varepsilon_2$$

for all sufficiently large n.

Definition. Let F_n be a sequence of distributions on $\mathbb{R}^k$ and F_0 be a given distribution on $\mathbb{R}^k$. We say that $F_n \xrightarrow{\mathcal{L}} F_0$ iff

$$\int_{\mathbb{R}^k} b(x)dF_n(x) \to \int_{\mathbb{R}^k} b(x)dF_\theta(x)$$

for all bounded continuous functions $b : \mathbb{R}^k \to \mathbb{R}^1$.

4 (Hájek). Let $F_{n,\theta} = \mathcal{L}\big(\sqrt{n}(\tau_n - \theta)\big)$ and suppose that $F_{n,\theta+\frac{a}{\sqrt{n}}} \xrightarrow{\mathcal{L}} G$ for all $|a| \leq 1$. Then G is the distribution function of $X+Y$, where X is $N\big(0, 1/I_1(\theta)\big)$ and X and Y are independent. (This is true for all θ. G can depend on θ.)

Corollary. *The variance of G (if it exists) is at least $1/I_1(\theta)$.*

Conclusion. At least in the iid case, Fisher's assertion is essentially correct.

Proof of (3) (outline). Choose $\theta \in \Theta$ and $a > 0$, and let $\delta_n = \theta + \frac{a}{\sqrt{n}}$. For fixed n, consider testing θ against δ_n. $\frac{\ell(\delta_n|s_n)}{\ell(\theta|s_n)}$ is the optimal (LR) test statistic, whose logarithm is

$$L_n(\delta_n) - L(\theta) = \frac{a}{\sqrt{n}} L'(\theta) + \frac{a^2}{2n} L''(\theta_n^*) = \frac{a}{\sqrt{n}} L'(\theta) - \frac{1}{2} a^2 I_1(\theta) + \cdots,$$

where the omitted terms are negligible. Let

$$K_n(s_n) = \frac{1}{\sqrt{a^2 I_1(\theta)}} \Big(L(\delta_n \mid s_n) - L(\theta \mid s_n) + \frac{1}{2} a^2 I_1(\theta) \Big).$$

K_n is equivalent to the LR statistic and $K_n \xrightarrow{\mathcal{L}} N(0,1)$ under P_θ. Consider the distribution of K_n under δ_n,

$$\begin{aligned} P_{\delta_n}(K_n < z) &= \int_{K_n < z} dP_{\delta_n}^{(n)} = \int_{K_n(s_n)<z} e^{L(\delta_n|s_n) - L(\theta|s_n)} dP_\theta^{(n)}(s_n) \\ &= \int_{K_n(s_n)<z} e^{-\frac{1}{2}a^2 I_1(\theta) + \sqrt{a^2 I_1(\theta)} K_n(s_n)} dP_\theta^{(n)}(s_n) = \int_{y<z} e^{-\frac{1}{2}a^2 I_1(\theta) + \sqrt{a^2 I_1(\theta)} y} dF_n(y) \\ &\to \int_{y<z} e^{-\frac{1}{2}a^2 I_1(\theta) + \sqrt{a^2 I_1(\theta)} y} d\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}a^2 I_1(\theta) + \sqrt{a^2 I_1(\theta)} y - \frac{1}{2} y^2} dy \\ &= P\big(N(0,1) < z - \sqrt{a_2 I_1(\theta)}\big), \end{aligned}$$

where $F_n(y) = P_\theta(K_n < y)$. Note that $F_n(y) \to \Phi(y)$.

Given a sequence $\{t_n\}$ such that $\overline{\lim}_{n\to\infty} P_{\delta_n}(t_n \geq \delta_n) \geq 1/2$, choose $z > \sqrt{a^2 I_1(\theta)}$. Then, by the above result, $P_{\delta_n}(K_n \geq z) < 1/2$ for all sufficiently large n. Regard $\{t_n \geq \delta_n\}$ and $\{K_n \geq z_n\}$ as critical regions for the test; then, by the Neyman-Pearson lemma, we have that, for some subsequence $\{n_k\}$, $P_\theta(K_{n_k} > z) \leq P_\theta(t_{n_k} \geq \delta_{n_k})$ for all sufficiently large k; but

$$P_\theta(t_n \geq \delta_n) = P_\theta\big(\sqrt{n}(t_n - \theta) \geq a\big) \quad \text{and} \quad P_\theta(K_n \geq z) \to 1 - \Phi(z),$$

so

$$z > \sqrt{a^2 I_1(\theta)} \Rightarrow P_\theta(K_{n_k} > z) < P_\theta(t_{n_k} \geq \theta + a/\sqrt{n_k}).$$

Letting $k \to \infty$, we find that

$$P\big(N(0,1) \geq z\big) \leq P\big(N(0,1) \geq a/\sqrt{v(\theta)}\big)$$

and hence $z > a/\sqrt{v(\theta)}$. Since z was arbitrary, we must have $\sqrt{a^2 I_1(\theta)} \geq a/\sqrt{v(\theta)}$ and hence $v(\theta) \geq 1/I_1(\theta)$. $\square$

Lecture 30

Proof of (2). Assume only (***), i.e., that $\sqrt{n}(t_n - \theta) \xrightarrow{\mathcal{L}_\theta} N\big(0, v(\theta)\big)$ for $\theta \in \Theta$, and let J be a bounded subinterval of Θ, say (a, b). Let

$$\Psi_n(\theta) = P_\theta(t_n > \theta) \quad \text{and} \quad \varphi_n(\theta) = \left|\Psi_n(\theta) - \frac{1}{2}\right|.$$

Then $0 \le \varphi_n(\theta) \le \frac{1}{2}$ and, from (***), $\Psi_n(\theta) \to \frac{1}{2}$ and $\varphi_n(\theta) \to 0$ for each θ. Hence $\theta \mapsto I_J(\theta)\varphi_n(\theta)$, where I_J is an indicator function, is bounded on Θ and tends to 0, so $\int_\Theta I_J(\theta)\varphi_n(\theta)d\theta \to 0$, or

$$\int_{\mathbb{R}^1} I_J\left(\delta + \frac{1}{\sqrt{n}}\right)\varphi_n\left(\delta + \frac{1}{\sqrt{n}}\right)d\delta \to 0;$$

but $I_J(\delta + \frac{1}{\sqrt{n}}) \to I_J(\delta)$ except for δ an endpoint of J, so

$$\int_{\mathbb{R}^1} I_J(\delta)\varphi_n\left(\delta + \frac{1}{\sqrt{n}}\right)d\delta \to 0.$$

Noticing that $I_J(\delta)\varphi_n\big(\delta+\frac{1}{\sqrt{n}}\big) \ge 0$, we have $I_J(\delta)\varphi_n\big(\delta+\frac{1}{\sqrt{n}}\big) \to 0$ in Lebesgue measure, so that there is some sequence $\{n_k\}$ such that $I_J(\delta)\varphi_{n_k}\big(\delta + \frac{1}{\sqrt{n_k}}\big) \to 0$ a.e.(Lebesgue); thus $\varphi_{n_k}\big(\delta+\frac{1}{\sqrt{n_k}}\big) \to 0$ a.e.(Lebesgue) on J – i.e., $P_{\theta+\frac{1}{\sqrt{n_k}}}\big(t_{n_k} > \theta+\frac{1}{\sqrt{n_k}}\big)-\frac{1}{2} \to 0$ a.e. on J. Returning to the original sequence, we have that $\overline{\lim}_{n\to\infty} P_{\theta+\frac{1}{\sqrt{n}}}\big(t_n > \theta+\frac{1}{\sqrt{n}}\big) \ge 1/2$ a.e. on J and so, from (3), $v(\theta) \ge 1/I_1(\theta)$ a.e. on J. Since J was *any* bounded subinterval of Θ, this means that $v(\theta) \ge 1/I_1(\theta)$ a.e. on Θ. □

General regular case

For each n, let $(S_n, \mathcal{A}_n, P_\theta^{(n)})$ be an experiment with common parameter

$$\theta = (\theta_1, \ldots, \theta_p) \in \Theta,$$

where Θ is open in $\mathbb{R}^p$, such that S_n consists of points s_n. No relation between n and $n+1$ is assumed.

In Examples 1–5, we have $S_n = \underbrace{X \times \cdots \times X}_{n \text{ times}}$ and $P_\theta^{(n)} = P_\theta \times \cdots P_\theta$. In Examples 6 and 7, $P_\theta^{(n)}$ is the distribution of $s_n = (X_1, \ldots, X_n)$, where the X_i are *not* iid.

Example 8. For $n = 2, 3, \ldots$, let n_1 and n_2 be positive integers such that $n = n_1 + n_2$. Let $s_n = (X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2})$, where $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ are independent, $X_1, \ldots, X_{n_1}$ are $N(\mu_1, \sigma^2)$ distributed and $Y_1, \ldots, Y_{n_2}$ are $N(\mu_2, \sigma^2)$ distributed. Here $\theta = (\mu_1, \mu_2, \sigma^2)$ is entirely unknown. This is a three-parameter exponential family, and the complete sufficient statistic is

$$\Big(\sum_{i=1}^{n_1} X_i, \sum_{i=1}^{n_2} Y_i, \sum_{i=1}^{n_1} X_i^2 + \sum_{i=1}^{n_2} Y_i^2\Big).$$

If $n_1/n_2 \to \rho$ as $n \to \infty$ for some $0 < \rho < \infty$, all regularity conditions to follow are satisfied.

The local asymptotic normality condition

Choose $\theta \in \Theta$ and assume that $dP_\delta^{(n)}(s_n) = \Omega_{\delta,\theta}(s_n)dP_\theta^{(n)}(s_n)$ holds for all δ in a neighborhood of θ.

Condition LAN (at $\theta \in \Theta$). For each $a \in \mathbb{R}^p$,

$$\log_e\big(\Omega_{\theta+\frac{a}{\sqrt{n}},\theta}(s_n)\big) = az_n'(\theta) - \frac{1}{2}a'I_1(\theta)a + \Delta_n(\theta, s_n),$$

where I_1 is a fixed $p \times p$ positive definite matrix, $z_n(\theta) \in \mathbb{R}^p$ and $z_n(\theta) \xrightarrow{\mathcal{L}_\theta} N\big(0, I_1(\theta)\big)$ and $\Delta_n(\theta, s_n) \to 0$ in $P_\theta^{(n)}$-probability.

Note.

i. If $s_n = (X_1, \ldots, X_n)$, where the X_is are iid, and I_1 is the information matrix for X_1, then LAN is satisfied for this I_1; but the LAN condition holds in some "irregular" cases also – see Example 1(b).

ii. The right-hand side in LAN with Δ_n omitted is exactly the log-likelihood in the multivariate normal translation-parameter case. See Example 4.

Let $g : \Theta \to \mathbb{R}^1$ be continuously differentiable and write $h(\theta) = \operatorname{grad} g(\theta)$.

2^p (Le Cam). If $t_n = t_n(s_n)$ is an estimate of g such that

$$\sqrt{n}\big(t_n - g(\theta)\big) \xrightarrow{\mathcal{L}_\theta} N\big(0, v(\theta)\big)\ \forall \theta \in \Theta,$$

then $\{\theta : v(\theta) < b_1(\theta)\}$ is of (p-dimensional) Lebesgue measure 0 if we let $b_1(\theta) = h(\theta)I_1^{-1}(\theta)h'(\theta)$.

4^p (Hájek). Suppose that $u_n : S_n \to \Theta$ is s.t.

$$\sqrt{n}\big(u_n - (\theta + a/\sqrt{n})\big) \xrightarrow{\mathcal{L}_{\theta+a/\sqrt{n}}} u_\theta$$

(u_θ independent of a), then u_θ may be represented as $v_\theta + w_\theta$, where v_θ and w_θ are independent and $v_\theta \sim N\big(0, I_1^{-1}(\theta)\big)$.

Note. No uniformity in a is needed in Hájek's theorem.

From the above we see that, for large n, the $N\big(0, I_1^{-1}(\theta)/n\big)$ distribution is nearly the best possible for estimates of θ. n is the "sample size", or cost of observing s_n.

Sufficient conditions for LAN

Suppose that $L(\theta \mid s_n)$ exists for each n, i.e., that $dP_\theta^{(n)}(s_n) = e^{L(\theta|s_n)} d\nu^{(n)}(s_n)$ for all n, and that, for each n, $L(\cdot \mid s_n)$ has at least two continuous derivatives. We write $\ell = e^L$. Let $L^{(1)}(\theta \mid s_n) = \operatorname{grad} L(\theta \mid s_n)$.

Condition 1. $\frac{1}{\sqrt{n}} L^{(1)}(\theta \mid s_n) \xrightarrow{\mathcal{L}_\theta} N\big(0, I_1(\theta)\big)$ for some positive definite I_1.

Condition 2. $\frac{1}{n}\{L_{ij}(\theta \mid s_n)\} \to -I_1(\theta)$ in $P_\theta^{(n)}$-probability.

Condition 3. With

$$M(\theta, \gamma, s_n) := \frac{1}{n} \max_{\substack{\|\delta-\theta\|\le\gamma \\ i,j=1,\dots,p}} \{|L_{ij}(\delta \mid s_n) - L_{ij}(\theta \mid s_n)|\},$$

$\lim_{r\downarrow 0} \overline{\lim}_{n\to\infty} P_\theta^{(n)}\big(M(\theta,\gamma,s_n) > \varepsilon\big) = 0$ for every $\varepsilon > 0$.

Conditions 1–3 imply LAN with $\Delta_n \to 0$, and also the following:

1^p (Fisher). Under Conditions 1–3, if $\hat\theta_n = \hat\theta_n(s_n)$, the MLE of θ, exists and is consistent, then

$$\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{\mathcal{L}_\theta} N\big(0, I_1^{-1}(\theta)\big) \ \forall\theta \in \Theta.$$

Definition. Let $u_n = u_n(s_n)$ be an estimate of θ. u_n is CONSISTENT if $u_n \xrightarrow{P_\theta} \theta$ for all θ, or, equivalently, $(u_n - \theta)(u_n - \theta)' \xrightarrow{P_\theta} 0$. u_n is $\sqrt{n}$-CONSISTENT if $n(u_n - \theta)(u_n - \theta)'$ is bounded in P_θ for all θ. (We say that Y_n is BOUNDED in P if, given any $\varepsilon > 0$, we may find k such that $P(|Y_n| > k) \le \varepsilon$ for all n sufficiently large.)

1^p (continued). If u_n is a $\sqrt{n}$-consistent estimate of θ and

$$u_n^* = u_n + \big\{(L_{ij}(\theta \mid s_n))^{-1} L^{(1)}(\theta \mid s_n)\big|_{\theta=u_n}\big\}$$

and

$$u_n^{**} = u_n + \big\{I_n(\hat\theta_n)^{-1} L^{(1)}(\theta \mid s_n)\big|_{\theta=u_n}\big\},$$

then u_n^* and u_n^{**} are both $AN\big(\theta, I_1^{-1}(\theta)/n\big)$. Consequently, $t_n^* = g(u_n^*)$ and $t_n^{**} = g(u_n^{**})$ are both $AN\big(g(\theta), b_1(\theta)/n\big)$, where $b_1(\theta) = h(\theta) I_1^{-1}(\theta) h'(\theta)$.