Continuous Mapping

Theorem 1.1.1. Suppose (X_n) is a sequence of random variables and g is a continuous mapping, then:

i. If $X_n \xrightarrow{as} X$, then $g(X_n) \xrightarrow{as} g(X)$.

ii. If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.

iii. If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.

□

Proof. (**i.**) Suppose $X_n \xrightarrow{as} X$, then there exits a set $A \subseteq \Omega$ such that:

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for $\forall \omega \in A$ and $\mathbb{P}(A) = 1$. Since g is a continuous function, on the same set:

$$g\big\{X(\omega)\big\} = g\left\{\lim_{n\to\infty} X(\omega)\right\} = \lim_{n\to\infty} g\big\{X(\omega)\big\}.$$

Conclude that $g(X_n) \xrightarrow{as} g(X)$.

(**ii.**) Suppose $X_n \xrightarrow{p} X$, and let $m \in \mathbb{N}$. By total probability:

$$\begin{aligned}\mathbb{P}\big(|g(X_n) - g(X)| > \epsilon\big) =& \mathbb{P}\big(|g(X_n) - g(X)| > \epsilon, |X_n - X| < m^{-1}\big) \\ &+ \mathbb{P}\big(|g(X_n) - g(X)| > \epsilon, |X_n - X| > m^{-1}\big).\end{aligned}$$

Proposition 1.1.1.

$$\lim_{m\to\infty}\lim_{n\to\infty} \mathbb{P}\big(|g(X_n) - g(X)| > \epsilon, |X_n - X| < m^{-1}\big) = 0.$$

♦

Define:

$$A_{mn} = \big\{\omega : |g(X_n) - g(X)| > \epsilon, |X_n - X| < m^{-1}\big\}, \qquad A_m = \bigcup_{n=1}^{\infty} A_{mn}.$$

Observe first that the sequence (A_m) is decreasing, since if $m_2 > m_1$ and $\omega \in A_{m_2}$ then $\omega \in A_{m_1}$. Fix ω, and hence $X(\omega)$. By continuity of g at $X(\omega)$, there exists δ such that if $|\xi - X(\omega)| < \delta$ then $|g(\xi) - X(\omega)| < \epsilon$. Chose M such that $M^{-1} < \delta$, then $\omega \notin A_m$ for $m \geq M$. Since ω was arbitrary:

$$\bigcap_{m=1}^{\infty} A_m = \emptyset.$$

Consequently,

$$\lim_{m\to\infty}\lim_{n\to\infty} \mathbb{P}(A_{mn}) \leq \lim_{m\to\infty} \mathbb{P}(A_m) = \mathbb{P}\left(\bigcap_{m=1}^{\infty} A_m\right) = 0.$$

Proposition 1.1.2.

$$\lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}\big(|g(X_n)-g(X)|>\epsilon, |X_n-X|>m^{-1}\big)=0.$$

♦

This follows from $X_n \xrightarrow{p} X$ since:

$$\lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}\big(|g(X_n)-g(X)|>\epsilon, |X_n-X|>m^{-1}\big)$$
$$\leq \lim_{m\to\infty}\Big\{\lim_{n\to\infty}\mathbb{P}\big(|X_n-X|>m^{-1}\big)\Big\}=\lim_{m\to\infty}\big\{0\big\}=0.$$

Overall, if $X_n \xrightarrow{p} X$, then for $\forall\epsilon>0$:

$$\lim_{n\to\infty}\mathbb{P}\big(|g(X_n)-g(X)|>\epsilon\big)=0.$$

(**iii.**) Suppose $X_n \xrightarrow{d} X$. Let F_n denote the distribution of X_n and F the distribution of X. Also, let G_n denote the distribution of $g(X_n)$ and G the distribution of $g(X)$. The aim is to show that $G_n \to G$ (at points of continuity of G). By the Skorokhod representation theorem, there exist (ξ_n) and ξ, defined on a common probability space, such that ξ_n has distribution F_n, ξ has distribution F, and $\xi_n \xrightarrow{as} \xi$. Now, by (i.) $g(\xi_n) \xrightarrow{as} g(\xi)$, and by the convergence hierarchy, $g(\xi_n) \xrightarrow{as} g(\xi)$ implies $g(\xi_n) \xrightarrow{d} g(\xi)$. Since $g(\xi_n)$ has distribution G_n and $g(\xi)$ has distribution G, conclude $G_n \to G$ (at points of continuity of G). ■

Slutsky's

Remark 1.2.1. Recall from the portmanteau theorem that the following statements are equivalent:

i. $X_n \xrightarrow{d} X$,

ii. $E\big\{g(X_n)\big\} \to E\big\{g(X)\big\}$ for every bounded, continuous function g.

iii. $E\big\{\mathcal{L}(X_n)\big\} \to E\big\{\mathcal{L}(X)\big\}$ for every bounded, Lipschitz function $\mathcal{L}$.

♦

Proposition 1.2.1. If $X_n \xrightarrow{d} X$ and $||Y_n - X_n|| = o_p(1)$, then $Y_n \xrightarrow{d} X$. ♦

Proof. Let $\mathcal{L}$ denote a bounded, Lipschitz function. Then, there $\exists$ a constant L such that $||\mathcal{L}(x) - \mathcal{L}(y)|| \leq L||x - y||$. Moreover, since $\mathcal{L}$ is bounded, $||\mathcal{L}(x)|| \leq M$. First consider:

$$\begin{aligned} ||E\big\{\mathcal{L}(Y_n)\big\} - E\big\{\mathcal{L}(X_n)\big\}|| \leq & E||\mathcal{L}(Y_n) - \mathcal{L}(X_n)|| \\ = & E\left\{||\mathcal{L}(Y_n) - \mathcal{L}(X_n)||I(||Y_n - X_n|| < \epsilon)\right\} \\ & + E\left\{||\mathcal{L}(Y_n) - \mathcal{L}(X_n)||I(||Y_n - X_n|| \geq \epsilon)\right\} \\ \leq & LE\left\{||Y_n - X_n||I(||Y_n - X_n|| < \epsilon)\right\} + 2M\mathbb{P}\big\{||Y_n - X_n|| \geq \epsilon\big\} \\ < & L\epsilon\mathbb{P}\big\{||Y_n - X_n|| < \epsilon\big\} + 2M\mathbb{P}\big\{||Y_n - X_n|| \geq \epsilon\big\}. \end{aligned}$$

Next, consider:

$$\begin{aligned} ||E\big\{\mathcal{L}(Y_n)\big\} - E\big\{\mathcal{L}(X)\big\}|| \leq & ||E\big\{\mathcal{L}(Y_n)\big\} - E\big\{\mathcal{L}(X_n)\big\}|| + ||E\big\{\mathcal{L}(X_n)\big\} - E\big\{\mathcal{L}(X)\big\}|| \\ < & L\epsilon + 2M\mathbb{P}\big\{||Y_n - X_n|| \geq \epsilon\big\} + ||E\big\{\mathcal{L}(X_n)\big\} - E\big\{\mathcal{L}(X)\big\}||. \end{aligned}$$

Taking the limit as $n \to \infty$, the second and third terms vanish since $||Y_n - X_n|| = o_p(1)$ and $X_n \xrightarrow{d} X$, hence:

$$\lim_{n\to\infty} ||E\big\{\mathcal{L}(Y_n)\big\} - E\big\{\mathcal{L}(X)\big\}|| < K\epsilon.$$

Since $\epsilon > 0$ was arbitrary, the limit vanishes. Conclude that $Y_n \xrightarrow{d} X$. ■

Proposition 1.2.2. Suppose $X_n \xrightarrow{d} X$ and Y_n converges in probability to a constant α ($Y_n \xrightarrow{p} \alpha$), then (X_n, Y_n) converge jointly in distribution to (X, α). ♦

Proof. Let $g(x, y)$ denote a bounded, continuous function. Define $h(x) = f(x, \alpha)$. Since $X_n \xrightarrow{d} X$, by the portmanteau theorem, $E\big\{h(X_n)\big\} \to E\big\{h(X)\big\}$, and equivalently:

$$E\big\{g(X_n, \alpha)\big\} \to E\big\{g(X, \alpha)\big\}.$$

Thus $(X_n, \alpha) \xrightarrow{d} (X, \alpha)$. Now $||(X_n, Y_n) - (X_n, \alpha)|| = ||Y_n - \alpha|| = o_p(1)$. Together, $(X_n, \alpha) \xrightarrow{d} (X, \alpha)$ and $||(X_n, Y_n) - (X_n, \alpha)|| = o_p(1)$ imply $(X_n, Y_n) \xrightarrow{d} (X_n, \alpha)$. ■

Theorem 1.2.1 (Slutsky's). If $X_n \xrightarrow{d} X$ and Y_n converges in probability to a constant α ($Y_n \xrightarrow{p} \alpha$), then for a continuous mapping f:

$$f(X_n, Y_n) \xrightarrow{d} f(X, \alpha).$$

□

Proof. By the preceding propositions, $(X_n, Y_n) \xrightarrow{d} (X, \alpha)$, and by the continuous mapping theorem $f(X_n, Y_n) \xrightarrow{d} f(X, \alpha)$. ■

Law of Large Numbers

Theorem 1.3.1 (LLN). Suppose (X_i) is a sequence of independent random variables. If the following conditions hold:

i. $\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{P}\big(|X_i|>n\big)=0$,

ii. and $\lim_{n\to\infty}n^{-2}\sum_{i=1}^{n}E\big\{X_i^2 I(|X_i|\le n)\big\}=0$,

then,

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\big(X_i-\mu_{n,i}\big)=0,$$

where $\mu_{n,i}=E\big\{X_i I(|X_i|\le n)\big\}$. □

Proof. Define $Y_{n,i}=X_i I(|X_i|\le n)$, and:

$$S_n=\sum_{i=1}^{n}X_i, \qquad T_n=\sum_{i=1}^{n}Y_{n,i}.$$

Proposition 1.3.1. The original sum S_n and the truncated sum T_n converge to the same limit: $S_n-T_n=o_p(1)$. ♦

This follows from hypothesis (i.) since:

$$\begin{aligned}\mathbb{P}\big(|S_n-T_n|>\epsilon\big)&\le\mathbb{P}\big(S_n\ne T_n\big)\le\mathbb{P}\left(\bigcup_{i=1}^{n}\{X_i\ne Y_{n,i}\}\right)\\&\le\sum_{i=1}^{n}\mathbb{P}\big(X_i\ne Y_i\big)=\sum_{i=1}^{n}\mathbb{P}\big(|X_i|>n\big)\to 0.\end{aligned}$$

Proposition 1.3.2.

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\big(Y_{n,i}-\mu_{n,i}\big)=0.$$

♦

This follows from hypothesis (ii.) since:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\big(Y_{n,i}-\mu_{n,i}\big)\right|\ge\epsilon\right)&\le\frac{1}{\epsilon^2n^2}\text{Var}\big(T_n\big)\le\frac{1}{\epsilon^2n^2}\sum_{i=1}^{n}E\big(Y_{n,i}^2\big)\\&\le\frac{1}{\epsilon^2n^2}\sum_{i=1}^{n}E\big\{X_i^2I(|X_i|\le n)\big\}=0.\end{aligned}$$

Overall:

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\big(X_i-\mu_{n,i}\big)=\underset{n\to\infty}{\text{plim}}\frac{S_n-T_n}{n}+\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\big(Y_{n,i}-\mu_{n,i}\big)=0.$$

■

Example 1.3.1. Let (X_i) denote a sequence of independent random variables, with mean $E(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$. Suppose:

$$\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \mu_i$$

exists, and that the variances are uniformly bounded $\text{Var}(X_i) \leq M$. Then:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu.$$

To see this, let $Y_i = X_i - \mu_i$ such that $E(Y_i) = 0$ for $\forall i \in \mathbb{N}$, then:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}\big(|Y_i| > n\big) \leq \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\text{Var}(Y_i)}{n^2} \leq \lim_{n\to\infty} \frac{n\cdot M}{n^2} = 0.$$

The second condition follows similarly:

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} E\big\{|Y_i| I(|Y_i| \leq n)\big\} \leq \lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(Y_i) = \lim_{n\to\infty} \frac{n\cdot M}{n^2} = 0.$$

♠

Theorem 1.3.2 (Kolmogorov LLN). Suppose (X_n) are IID. There exists a constant α such that:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{as} \alpha$$

if and only if $E(X_i) < \infty$, in which case $\alpha = E(X_i)$. □

Remark 1.3.1. See Resnick (2014), theorem 7.5.1; Serfling (1980), section 1.8. ♦

3.1 Glivenko-Cantelli

Theorem 1.3.3. Suppose (X_i) are IID random variables with common distribution function F. Let $\mathbb{F}_n$ denote the empirical distribution function:

$$\mathbb{F}_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} I\big\{X_i(\omega) \leq x\big\}.$$

Then $\mathbb{F}_n$ converges to F uniformly almost surely:

$$\sup_{x\in\mathbb{R}} \big|\mathbb{F}_n(x,\omega) - F(x)\big| \xrightarrow{as} 0.$$

□

Remark 1.3.2. See Resnick (2014), theorem 7.5.2. ♦

Central Limit Theorem

4.1 IID Case

Proposition 1.4.1 (**Lindeberg-Levy CLT**). If (Y_i) are IID random variables with $E(Y_i) = 0$ and $E(Y_i^2) = 1$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{d} N\big(0,1\big).$$

♦

Proof. Consider the characteristic function of $n^{-1/2}\sum_{i=1}^{n} Y_i$:

$$\phi_n(t) = E(e^{it\cdot n^{-1/2}\sum_{i=1}^{n} Y_i}) = \big\{\phi(n^{-1/2}t)\big\}^n = \left\{\phi(0) + \frac{t}{n^{1/2}}\dot{\phi}(0) + \frac{t^2}{2n}\ddot{\phi}(0) + o(n^{-1})\right\}^n,$$

where $\phi(\cdot)$ is the characteristic function of Y. Now:

$$\dot{\phi}(0) = iE(Y) = 0, \qquad \ddot{\phi}(0) = -E(Y^2) = -1.$$

Thus:

$$\phi_n(t) = \left\{\phi(0) - \frac{t^2}{2n} + o(n^{-1})\right\}^n \to e^{-t^2/2}.$$

Since $\exp(-t^2/2)$ is the characteristic function of the standard normal distribution, by Levy's continuity theorem, $n^{-1/2}\sum_{i=1}^{n} Y_i$ converges in distribution to $N(0,1)$. ■

Theorem 1.4.1 (**Berry Essen**). Suppose (X_i) are IID with mean $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 > 0$ and $E|X_i|^3 < \infty$. Let F_n denote the finite-sample distribution of the normalized sum:

$$Z_n = \sqrt{n}\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma},$$

then there exists a constant $C > 0$ such that for $\forall n \in \mathbb{N}$:

$$\sup_{x\in\mathbb{R}}\big|F_n(x) - \Phi(x)\big| \leq \frac{CE|X_i - \mu|^3}{\sigma^3\sqrt{n}},$$

where Φ is the standard normal distribution function. □

Remark 1.4.1. See Serfling (1980), section 1.9. ♦

4.2 Lindeberg Feller

Theorem 1.4.2 (Lindeberg-Feller CLT). Let (X_i) denote a sequence of independent random variables. Define the centered random variables $Y_i = X_i - E(X_i)$. Suppose $\text{Var}(X_i) = E(Y_i^2) = \sigma_i^2 < \infty$. Let $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If, for $\forall \epsilon > 0$, the **Lindeberg condition** holds:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^n E\big\{X_i^2 I(|X_i| > \epsilon s_n)\big\} = 0, \tag{A_1}$$

then:

$$\frac{\sum_{i=1}^n \big(X_i - \mu_i\big)}{\sqrt{\big(\sum_{i=1}^n \sigma_i^2\big)}} \xrightarrow{d} N(0,1). \tag{B}$$

□

Remark 1.4.2. See Resnick (2014), theorem 9.8.1. For the multivariate extension, see Serfling (1980), section 1.9. ♦

Proposition 1.4.2. The Lindeberg condition (A_1) implies:

$$\lim_{n\to\infty} \frac{\max_{1\le i\le n} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2} = 0. \tag{A_2}$$

♦

Proof.

$$\begin{aligned}\frac{\sigma_i^2}{s_n^2} = \frac{E(Y_i^2)}{s_n^2} &= \frac{1}{s_n^2} E\big\{Y_i^2 I(|Y_i| \le \epsilon s_n)\big\} + \frac{1}{s_n^2} E\big\{Y_i^2 I(|Y_i| \le \epsilon s_n)\big\} \\ &\le \epsilon^2 + \frac{1}{s_n^2} \sum_{i=1}^n E\big\{Y_i^2 I(|Y_i| \le \epsilon s_n)\big\}.\end{aligned}$$

Since this bound is independent of i:

$$\frac{\max_{1\le i\le n} \sigma_i^2}{s_n^2} \le \epsilon^2 + \frac{1}{s_n^2} \sum_{i=1}^n E\big\{Y_i^2 I(|Y_i| \le \epsilon s_n)\big\}.$$

Taking the limit as $n \to \infty$, under (A_1):

$$\lim_{n\to\infty} \frac{\max_{1\le i\le n} \sigma_i^2}{s_n^2} \le \epsilon^2.$$

■

Proposition 1.4.3. Condition (A_2) implies *uniform asymptotic negligibility*:

$$\lim_{n\to\infty}\max_{1\le i\le n}\mathbb{P}\big(|Y_i|>\epsilon s_n\big)=0. \tag{A_3}$$

♦

Proof. By Chebyshev's inequality:

$$\lim_{n\to\infty}\max_{1\le i\le n}\mathbb{P}\big(|Y_i|>\epsilon s_n\big)\le\lim_{n\to\infty}\max_{1\le i\le n}\frac{E(Y_i^2)}{\epsilon^2 s_n^2}=\frac{1}{\epsilon^2}\lim_{n\to\infty}\frac{\max_{1\le i\le n}\sigma_i^2}{s_n^2}=0.$$

■

Proposition 1.4.4. The **Lyapunov condition** requires that $\exists\delta>0$ such that:

$$\lim_{n\to\infty}\frac{\sum_{i=1}^n E|X_i|^{2+\delta}}{s_n^{2+\delta}}=0. \tag{A_0}$$

The Lyapunov condition implies the *Lindeberg condition* (A_1). ♦

Proof.

$$\begin{aligned}\frac{1}{s_n^2}\sum_{i=1}^n E\big\{X_i^2 I(|X_i|>\epsilon s_n)\big\}&=\sum_{i=1}^n E\left\{\left|\frac{X_i}{s_n}\right|^2\cdot 1\cdot I(|X_i|/(\epsilon s_n)>1\right\}\\&\le\sum_{i=1}^n E\left\{\left|\frac{X_i}{s_n}\right|^2\left|\frac{X_i}{s_n}\right|^\delta I(|X_i|/(\epsilon s_n)>1\right\}\\&\le\frac{1}{\epsilon^\delta}\frac{\sum_{i=1}^n E|X_i|^{2+\delta}}{s_n^{2+\delta}}\to 0.\end{aligned}$$

■

Remark 1.4.3. Based on the preceding, the Lyapunov implies Lindeberg implies uniform asymptotic negligibility:

$$A_0\implies A_1\implies A_2\implies A_3.$$

The Lindeberg condition implies asymptotic normality:

$$A_0\implies A_1\implies B.$$

Feller's theorem states that, under condition (A_2), asymptotic normality holds if and only if the Lindeberg condition holds:

$$A_2\implies\big(A_1\iff B\big).$$

♦

4.3 Triangular Array

Theorem 1.4.3 (Triangular Array CLT). Suppose $(X_{n,i})$ is a sequence of independent random variables, $Y_{n,i} = X_{n,i} - E(X_{n,i})$, $\text{Var}(X_{n,i}) = E(Y_{n,i}^2) = \sigma_{n,i}^2 < \infty$, and $s_n^2 = \sum_{i=1}^{r_n} \sigma_{n,i}^2$. If, for $\forall \epsilon > 0$, the Lindeberg condition holds:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{r_n} E\big\{Y_{n,i}^2 I(|Y_{n,i}| > \epsilon s_n)\big\} = 0,$$

then:

$$\frac{\sum_{i=1}^{r_n} \big(X_{n,i} - \mu_{n,i}\big)}{\sqrt{\big(\sum_{i=1}^{r_n} \sigma_{n,i}^2\big)}} \xrightarrow{d} N(0,1).$$

□

Remark 1.4.4. See Billingsley (1995), theorem 27.2. ♦

4.4 Delta Method

Theorem 1.4.4. Suppose r_n is a sequence of positive constants such that $r_n \to \infty$ as $n \to \infty$, and that:

$$r_n\big(T_n - \theta\big) \xrightarrow{d} X.$$

If g is continuously differentiable in a neighborhood of θ, then:

$$r_n\big\{g(T_n) - g(\theta)\big\} \xrightarrow{d} \dot{g}(\theta)X.$$

□

Proof. Since $r_n\big(T_n - \theta\big)$ converges in distribution, $r_n\big(T_n - \theta\big) = \mathcal{O}_p(1)$ or:

$$T_n - \theta = \mathcal{O}_p(r_n^{-1}).$$

Since $r_n \to \infty$, this shows $T_n = \theta + o_p(1)$. Take the Taylor expansion of g about θ:

$$g(T_n) = g(\theta) + \dot{g}(\theta)(T_n - \theta) + \mathcal{O}_p(r_n^{-2}),$$
$$r_n\big\{g(T_n) - g(\theta)\big\} = \dot{g}(\theta)r_n(T_n - \theta) + \mathcal{O}_p(r_n^{-1}).$$

By Slutsky's theorem:

$$r_n\big\{g(T_n) - g(\theta)\big\} \xrightarrow{d} \dot{g}(\theta)X.$$

■

Example 1.4.1. Suppose that:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \xrightarrow{d} N\big(0, \boldsymbol{\Sigma}\big).$$

Let $g : \mathbb{R}^p \to \mathbb{R}$ denote a continuous differentiable mapping, then:

$$\sqrt{n}\big\{g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}_0)\big\} \xrightarrow{d} N\big\{0, \dot{g}(\boldsymbol{\theta}_0)'\boldsymbol{\Sigma}\dot{g}(\boldsymbol{\theta}_0)\big\},$$

where:

$$\dot{g}(\boldsymbol{\theta}_0) = \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

♠

Remark 1.4.5. See also van der Vaart (1998), theorem 3.1. ♦