STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Apr 7

Lecturer: Xianyang Zhang

1 Convergence rate

1.1 Notation

Let us define

$$\begin{aligned}
&M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_\theta(X_i) = \mathbb{P}_n l_\theta, \\
&\hat{\theta}_n = \arg\min_{\theta\in\Theta} M_n(\theta), \\
&M(\theta) = \mathbb{P} l_\theta, \\
&\Delta_n(\theta) = (M_n(\theta) - M(\theta)) - (M_n(\theta_0) - M(\theta_0)) = (\mathbb{P}_n - \mathbb{P})(l_\theta - l_{\theta_0}).
\end{aligned}$$

Also, define the modulus of continuity

$$\mathbb{W}_n(\delta) = \sup_{d(\theta,\theta_0)\le\delta} |\Delta_n(\theta)|.$$

Let $\mathcal{F} = \{l_\theta - l_{\theta_0} : \theta \in \Theta\}$. Then

$$\Delta_n(\theta) = (\mathbb{P}_n - \mathbb{P}) f_\theta$$

where $f_\theta = l_\theta - l_{\theta_0} \in \mathcal{F}$.

1.2 Theorem

Suppose

$$M(\theta) \ge M(\theta_0) + \lambda d^2(\theta, \theta_0)$$

for θ near θ_0 for some $\lambda > 0$. Let ϕ be a function such that $\phi(\delta) \le c\delta^\alpha$ for $\alpha \in (0, 2)$ and some constant $c > 0$. Assume,

$$\mathbb{E}[\mathbb{W}_n(\delta)] \le \frac{\phi(\delta)}{\sqrt{n}}.$$

Let $r_n \to \infty$ such that $c r_n^{2-\alpha} \le \sqrt{n}$. If $\hat{\theta}_n \xrightarrow{p} \theta_0$, then

$$r_n d(\hat{\theta}_n, \theta_0) = O_p(1).$$

1.3 Proof

We aim to show that $r_n d(\hat{\theta}_n, \theta_0) = O_p(1)$. That is for any $\epsilon > 0$, there exists an t such that

$$\mathbb{P}(r_n d(\hat{\theta}_n, \theta_0) > 2^t) \le \epsilon.$$

Since $\hat{\theta}_n \xrightarrow{p} \theta_0$, for a fixed $\eta > 0$, we know that

$$\mathbb{P}[d(\hat{\theta}_n, \theta_0) > \eta] \to 0.$$

The key technique we employ here is the so-called peeling device. Define

$$S_{n,j} = \{\theta : 2^j < r_n d(\theta, \theta_0) \leq 2^{j+1}\}.$$

Then we have

$$\begin{aligned}
\mathbb{P}[r_n d(\hat{\theta}_n, \theta_0) > 2^t] &\leq \mathbb{P}[r_n d(\hat{\theta}_n, \theta_0) > 2^t, d(\hat{\theta}_n, \theta_0) \leq \eta] + \mathbb{P}[d(\hat{\theta}_n, \theta_0) > \eta] \\
&\leq \mathbb{P}[\exists j \geq t \text{ such that } 2^j \leq r_n \eta \text{ and } \hat{\theta}_n \in S_{n,j}] + o(1) \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \mathbb{P}[\hat{\theta}_n \in S_{n,j}] + o(1).
\end{aligned}$$

Since $\hat{\theta}_n \in S_{n,j}$, there exists $\theta \in S_{n,j}$ such that $M_n(\theta) \leq M_n(\theta_0)$. By the assumption, we have

$$\begin{aligned}
M_n(\theta_0) - M(\theta_0) &\geq M_n(\theta) - M(\theta) + M(\theta) - M(\theta_0) \\
&\geq M_n(\theta) - M(\theta) + \lambda[d(\theta, \theta_0)]^2.
\end{aligned}$$

which implies that

$$(M_n(\theta_0) - M(\theta_0)) - (M_n(\theta) - M(\theta)) \geq \lambda[d(\theta, \theta_0)]^2 > 0.$$

Thus we get

$$|\Delta_n(\theta)| \geq \lambda[d(\theta, \theta_0)]^2 > \lambda \frac{2^{2j}}{r_n^2}.$$

Using this fact, we have for any $\epsilon > 0$

$$\begin{aligned}
\sum_{j \geq t; 2^j \leq r_n \eta} \mathbb{P}[\hat{\theta}_n \in S_{n,j}] &\leq \sum_{j \geq t; 2^j \leq r_n \eta} \mathbb{P}\left(\exists \theta \in S_{n,j} \text{ such that } |\Delta_n(\theta)| > \lambda \frac{2^{2j}}{r_n^2}\right) \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \mathbb{P}\left(\sup_{\theta \in S_{n,j}} |\Delta_n(\theta)| > \lambda \frac{2^{2j}}{r_n^2}\right) \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \frac{\mathbb{E}\left[\sup_{\theta \in S_{n,j}} |\Delta_n(\theta)|\right]}{\lambda \frac{2^{2j}}{r_n^2}} \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \frac{r_n^2}{\lambda 4^j} \frac{\phi(\frac{2^{j+1}}{r_n})}{\sqrt{n}} \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \frac{c r_n^2}{\lambda \sqrt{n} 4^j} \left[\frac{2^{j+1}}{r_n}\right]^\alpha \\
&\leq \sum_{j \geq t; 2^j \leq r_n \eta} \frac{1}{\lambda 2^{(2-\alpha)j - \alpha}} \\
&\leq \epsilon \quad \text{take t large enough}
\end{aligned}$$

where to get the third inequality we have used the fact that $\theta \in S_{n,j}$ implies $d(\theta, \theta_0) \leq \frac{2^{j+1}}{r_n}$ and we have used the assumption that $c r_n^{2-\alpha} \leq \sqrt{n}$ to get the second inequality from the bottom.

2 Convergence in Distribution

2.1 Weak convergence: equivalent definition

$X_n \xrightarrow{d} X$ if and only if $E[f(X_n)] \to E[f(X)]$ for all bounded and (Lipschitz) continuous function f.

2.2 Tightness

Let $\mathcal{D}$ be a metric space. A random variable $X : \Omega \to \mathcal{D}$ is tight if there is a compact set $K \subset \mathcal{D}$ such that $P(X \in K) > 1 - \epsilon$.

2.3 Asymptotically tight

A sequence of $\mathcal{D}$-valued random variables $\{X_n\}$ is asymptotically tight if for all $\epsilon > 0$, there is a compact set $K \subset \mathcal{D}$ such that $\limsup_{n\to\infty} P(X_n \notin K^\delta) < \epsilon$, where

$$K^\delta = \{y \in D \ ; \ d(y, K) < \delta\}$$

for $\delta > 0$ and the metric d associated with the metric space $\mathcal{D}$.

2.4 Prokhorov's theorem

(a) If $X_n \xrightarrow{d} X$ and X is tight, then X_n is asymptotically tight.

(b) If X_n is asymptotically tight, then there exists a subsequence X_{n_k} and tight X such that $X_{n_k} \xrightarrow{d} X$.

2.5 L^∞ class of functions

For a compact set T, define $L^\infty(T)$ as the set of functions $f : T \to \mathbb{R}$ such that

$$\|f\|_\infty = \sup_{t\in T} |f(t)| < \infty.$$

2.6 Uniformly continuous functions

Suppose we have a metric ρ on $T \times T$. Define $UC(T, \rho)$ as the set of uniformly continuous functions $f : T \to \mathbb{R}$. We note that $UC(T, \rho) \subseteq L^\infty(T)$. Consider $\sqrt{n}\sup_{t\in\mathbb{R}} |F_n(t) - F_0(t)|$. Note that

$$\sqrt{n}(F_n(t) - F_0(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{1}\{X_i \le t\} - F_0(t)) .$$

Suppose that for any $t_1, t_2, \ldots, t_k$,

$$\sqrt{n}\,(F_n(t_1) - F_0(t_1), \ldots, F_n(t_k) - F_0(t_k)) \xrightarrow{d} (\mathcal{G}_{F_0}(t_1), \cdots, \mathcal{G}_{F_0}(t_k)).$$

Then,

$$\text{cov}(\mathcal{G}_{F_0}(t_i), \mathcal{G}_{F_0}(t_j)) = F_0(t_i \wedge t_j) - F_0(t_i)F_0(t_j).$$

We expect the empirical process $\sqrt{n}(F_n - F_0)$ converges in distribution to a Gaussian process $\mathcal{G}_{F_0}$ with zero mean and covariance function as in the preceding display.

2.7 F_0-Brownian bridge

The limit process $\mathcal{G}_{F_0}$ is known as an F_0-Brownian bridge. In particular, if F_0 is uniform on $[0, 1]$,

$$\text{cov}(\mathcal{G}_{F_0}(t_i), \mathcal{G}_{F_0}(t_j)) = t_i \wedge t_j - t_i t_j$$

and $\mathcal{G}_{F_0}$ is called a standard Brownian bridge. For standard Brownian bridge, we can write

$$\mathcal{G}_{F_0}(t) = B(t) - tB(1)$$

where B is a standard Brownian motion.

2.8 Asymptotically equicontinuous: ASEC

Let $X_n \in L^\infty(T)$. We say that X_n's are asymptotically equicontinuous if for all $\epsilon, \eta > 0$, there is a finite partition $T_1, \dots, T_k$ of T such that

$$\limsup_{n\to\infty} P\left(\max_{1\le i\le k} \sup_{s,t\in T_i} |X_{n,s} - X_{n,t}| \ge \epsilon\right) \le \eta.$$

2.9 Example

Let $Z_i \in \mathbb{R}^d$ and $Z_i \sim^{i.i.d} P$. Assume $E\|Z_1\|^2 < \infty$ and $E[Z_i] = 0$. Define

$$X_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i^\top t$$

and

$$T = \{t \in \mathbb{R}^d \;:\; \|t\| \le M\}.$$

Then for small enough δ,

$$\begin{aligned} P\left(\sup_{s,t\in T, \|s-t\|\le\delta} |X_{n,s} - X_{n,t}| > \epsilon\right) \le & P\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right\| \delta \ge \epsilon\right) \\ \le & \frac{\delta^2}{\epsilon^2} E\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right\|^2\right] = \frac{\delta^2}{\epsilon^2} E[\|Z_1\|^2] < \eta. \end{aligned}$$