Feynman Liang
STAT 210A
Lecture #23
11/10/2016

1 Review

Uniform Weak LLN

$$W_1, W_2, \ldots \text{ iid} \in C(\underbrace{K}_{\text{compact}}) \tag{1.1}$$

$$\mathbb{E}\|W_i\|_\infty < \infty \tag{1.2}$$

$$\implies \bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i \overset{\text{unif}}{\xrightarrow{\text{p}}} \mathbb{E}W_i \tag{1.3}$$

$$\text{i.e. } \|\bar{W}_n - \mathbb{E}W_1\|_\infty \xrightarrow{\text{p}} 0 \tag{1.4}$$

Theorem 1.1 (Keener 9.4)**.** *Let $G_n, n \geq 1$ be random functions in $C(K)$, K compact, and for some fixed $g \in C(K)$*

$$\|G_n - g\|_\infty \xrightarrow{p} 0 \tag{1.5}$$

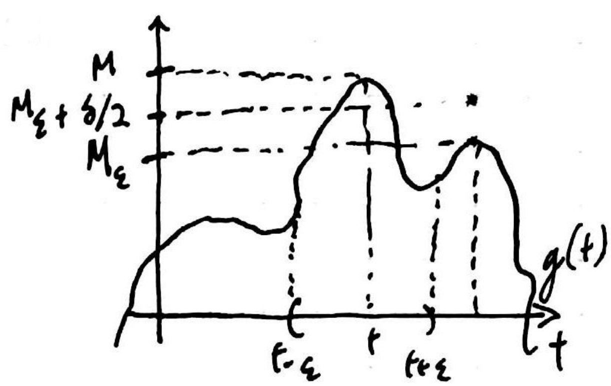
("G_n converging uniformly to g"), then

(a) If $t_n \xrightarrow{p} t^*$ (fixed), then $G_n(t_n) \xrightarrow{p} g(t^*)$

(b) If g has a unique maximizer $t^* \in K$, and $G_n(t_n) = \sup_{t\in K} G_n(t)$, then $t_n \xrightarrow{p} t^*$.

Proof. Item (a): Exercise

Item (b): Fix $\varepsilon > 0$, let $K_\varepsilon = K \setminus B_\varepsilon(t^*)$. Define $M = g(t^*) = \sup_{t\in K} g(t)$ and $M_\varepsilon = \sup_{t\in K_\varepsilon} g(t)$. Then $\delta = M - M_\varepsilon > 0$.



If $\|G_n - g\|_\infty \leq \delta/2$ (by $\xrightarrow{\text{p}}$, highly probable for suff large n), then

$$\sup_{t \in K_\varepsilon} G_n(t) < M_\varepsilon + \frac{\delta}{2} \tag{1.6}$$

$$\sup_{t \in K} G_n(t) \geq G_n(t^*) > M - \frac{\delta}{2} = M_\varepsilon + \delta/2 \tag{1.7}$$

$$P(t_n \in B_\varepsilon(t^*)) \geq P(\|G_n - g\|_\infty < \frac{\delta}{2}) \to 1 \tag{1.8}$$

□

Setup:

- $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta_0}(x)$
- $\theta_0 \Theta$
- $l_n(\theta; X) = \sum_{i=1}^n \log p_\theta(x_i)$
- Assume p_θ continuous in Θ, all p_θ distinct (model is *identifiable*)
- $\bar{W}_n = \frac{1}{n} l(\theta; X) - \frac{1}{n} l(\theta_0; X)$

Definition 1.2. The *Kullbeck-Leibler Divergence (KL-Divergence)* is

$$D_{KL}(\theta_0 \mid\mid \theta) = \mathbb{E}_{\theta_0} \log \frac{p_{\theta_0}(X)}{p_\theta(x)} = \mathbb{E}_{\theta_0}[-\bar{W}_n] \tag{1.9}$$

Lemma 1.3. *If $P_\theta \neq P_{\theta_0}$ then $D_{KL}(\theta_0 \mid\mid \theta) > 0$*

Proof. By Jensen's inequality

$$-D_{KL}(\theta_0 \mid\mid \theta) \leq \log \mathbb{E}_{\theta_0} \frac{p_\theta(X)}{p_{\theta_0}(X)} = \log \int_{x: p_{\theta_0} > 0} \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) dx \leq \log 1 = 0 \tag{1.10}$$

When $p_\theta \neq p_{\theta_0}$, Jensen's inequality is tight and the two have common support so the second inequality is also tight. □

Theorem 1.4 (Consistency of MLE if Θ compact)**.** *Let $W_i(\theta) = \log p_\theta(X_i) - \log p_{\theta_0}(X_i)$. If*

- *Θ compact*
- $\mathbb{E}_{\theta_0} \|W_1\|_\infty < \infty$
- *$p_\theta(x)$ cts in Θ (for a.e. x)*
- *$p_\theta \neq p_{\theta_0}$ for all $\theta \neq \theta_0$ (Identifiable)*

Then

$$\hat{\theta}_{MLE} \xrightarrow{p} \theta_0 \tag{1.11}$$

Proof. By definition $W_i(\theta) = \log \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)}$. By assumption p_θ continuous, we have $W_i \in C(\Theta)$ so

$$\|\bar{W}_n + D_{KL}(\theta_0 \mid\mid \theta)|_\infty \xrightarrow{\text{p}} 0 \tag{1.12}$$

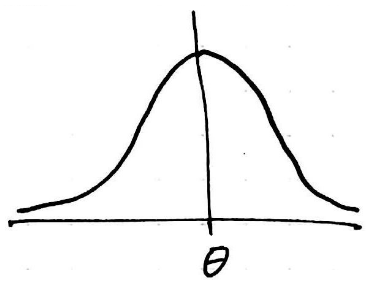
and we have uniform convergence. Also

$$\mathbb{E}_{\theta_0} W_i(\theta) = \underbrace{-D_{KL}(\theta_0 \mid\mid \theta)}_{\text{unique max at } \theta_0} \tag{1.13}$$

Applying theorem 1.1 yields $\hat{\theta}_{MLE} \xrightarrow{\text{p}} \theta_0$ □

Keener gives a weaker sufficient condition

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \log \frac{p_\theta(X)}{p_{\theta_0}(X)}\right] < \infty \tag{1.14}$$



What if Θ not compact?

Theorem 1.5. *Suppose*

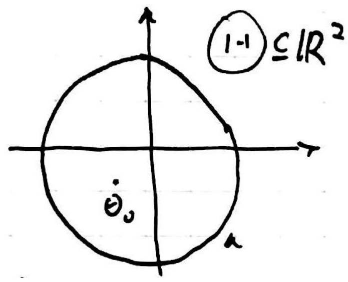
- $\Theta = \mathbb{R}^p$
- $p(\theta(x)$ *cts in* θ *(a.e.* x*)*
- $p_\theta \neq p_{\theta_0}$ $\forall \theta \neq \theta_0$ *(identifiable)*
- $p_\theta(x) \to 0$ *as* $\theta \to \infty$

If

(a) $\mathbb{E}_{\theta_0} \|1_K W_1\|_\infty < \infty$, $\forall$ *compact* $K \subset \Theta$

(b) $\mathbb{E}_{\theta_0} \sup_{\|\theta\| > a} W_1(\theta) < \infty$ *for some* $a > 0$

Then $\hat{\theta}_n \xrightarrow{p} \theta_0$



Proof.

$$
\begin{aligned}
& p_\theta(x) \xrightarrow{\text{a.s.}} 0 && \text{as } \theta \to \infty && (1.15)\\
\implies & \sup_{\|\theta\|>b} W_i(\theta) \to -\infty && \text{as } b \to \infty && (1.16)
\end{aligned}
$$

By dominated convergence

$$\mathbb{E}_{\theta_0}\left[\sup_{\|\theta\|>b} W_1(\theta)\right] \to -\infty \tag{1.17}$$

Choose $\delta > 0$ sufficiently large such that

$$\mathbb{E}_{\theta_0}\left[\sup_{\|\theta\|>b} W_1(\theta)\right] < -\delta < 0 \tag{1.18}$$

So

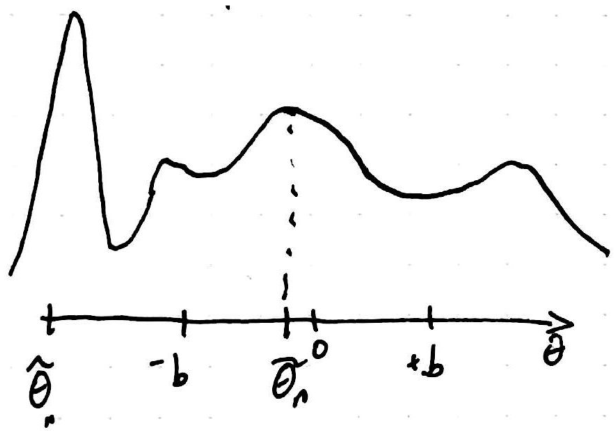
$$P_{\theta_0}\left(\sup_{\|\theta\|>b} \bar{W}_n(\theta) \geq -\delta\right) \to 0 \tag{1.19}$$

$$
\begin{aligned}
\sup_{\|\theta\|>b} \bar{W}_n(\theta) &= \frac{1}{n}\sup_{\|\theta\|>b}\sum_{i=1}^{n} W_i(\theta) && (1.20)\\
&= \frac{1}{n}\sum_{i=1}^{n}\sup_{\|\theta\|>b} W_i(\theta) && (1.21)\\
&\to \mathbb{E}_{\theta_0}[\sup \cdots] < \delta && (1.22)
\end{aligned}
$$

Let $\tilde{\theta}_n = \arg\max_{\|\theta\|\leq b} \overline{W}_n(\theta)$, which is consistent ($\xrightarrow{\text{P}} \theta_0$) because $\|\theta\| \leq b$ is a compact (closed-bounded) set.

$$
\begin{aligned}
P_{\theta_0}(\tilde{\theta}_n \neq \hat{\theta}_n) &= P_{\theta_0}\left(\sup_{\|\theta\|>b} \overline{W}_n(\theta) \geq \overline{W}_n(\theta_0)\right) && (1.23)\\
&\leq P_{\theta_0}\left(\sup_{\|\theta\|>b} \bar{W}_n(\theta) \geq -\delta\right) + P_{\theta_0}\left(\bar{W}_n(\theta_0) \leq -\delta\right) && (1.24)\\
&\to 0 \qquad \text{both by LLN} && (1.25)
\end{aligned}
$$

□



$\tilde{\theta}_n$
$-b$
$\bar{\theta}_n$
0
$+b$
θ