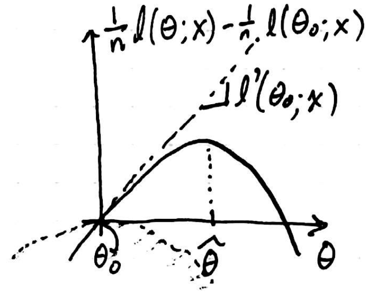
Feynman Liang
STAT 210A
Lecture #22
11/8/2016

1 Review



$$\hat{\theta} - \theta \approx \frac{l'(\theta_0; X)}{-l''(\theta_0; X)} \tag{1.1}$$

Applying mean value theorem

$$\hat{\theta} - \theta \approx \frac{l'(\theta_0; X)}{-l''(\tilde{\theta}; X)} \tag{1.2}$$

where $\tilde{\theta} \in [\theta_0, \theta_n]$.

Normalizing to get convergence in distribution

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{\frac{1}{\sqrt{n}} l'(\theta_0; X)}{-\frac{1}{n} l''(\tilde{\theta}; X)} \tag{1.3}$$

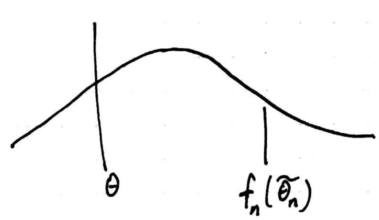
The numerator

$$\frac{1}{\sqrt{n}} l'(\theta_0; X) \Rightarrow N(0, J_1(\theta_0)) \tag{1.4}$$

Want the denominator

$$-\frac{1}{n} l''(\tilde{\theta}; X) \xrightarrow{p} -\mathbb{E} l''(\theta_0; X) = J_1(\theta_0) \tag{1.5}$$

which would then give asymptotic consistency: $\tilde{\theta}_n \xrightarrow{p} \theta_0$.



Today we will see what conditions are needed for asymptotic consistency.

2 Random Functions (Stochastic Processes)

$l(\theta; X)$ is a random function $\Theta \to \mathbb{R}$.

Define

$$\bar{W}_n = \frac{1}{n}\sum_{i=1}^{n}[l(\theta; X_i) - l(\theta_0; X_i)] \tag{2.1}$$

$$= \frac{1}{n}l(\theta; X) - \frac{1}{n}l(\theta_0; X) \tag{2.2}$$

$$= \frac{1}{n}\log\frac{\prod_i p_\theta(x_i)}{\prod_i p_{\theta_0}(x_i)} \tag{2.3}$$

For a compact set K, define $C(K) = \{f : K \to \mathbb{R}, f \text{ cts}\}$. Let $\|f\|_\infty = \sup|f(t)|$.

$$f_n \to f \text{ if } \|f_n - f\|_\infty \to 0 \tag{2.4}$$

$$f_n \xrightarrow{\text{p}} f \text{ if } \forall\varepsilon : P(\|f_n - f\|_\infty > \varepsilon) \to 0 \tag{2.5}$$

Lemma 2.1 (Keener 9.1)**.** *Let $W \in C(K)$ be a random function, with $\mathbb{E}\|W\|_\infty < \infty$. Then $\mathbb{E}W(t)$ is continuous in t and*

$$\sup_{t\in K}\mathbb{E}\left[\sup_{s:\|s-t\|<\varepsilon}|W(s) - W(t)|\right] \to 0 \tag{2.6}$$

as $\varepsilon \downarrow 0$.

Proof. See Keener. □

3 MLE is consistent

Step 1 $\|\bar{W}_n(\theta) - \mathbb{E}\bar{W}_n(\theta)\|_\infty \xrightarrow{\text{p}} 0$ for Θ compact

Step 2 $\mathbb{E}\bar{W}_n(\theta)$ maximized at $\theta = \theta_0$

Step 3 Conclude $\hat{\theta}_n \xrightarrow{\text{p}} \theta_0$ if Θ compact

Step 4 Handle $\Theta = \mathbb{R}^d$ not compact

Theorem 3.1 (Weak LLN for random functions)**.** *Let $W_1, W_2, \ldots$ iid random functions in $C(K)$, K compact.*

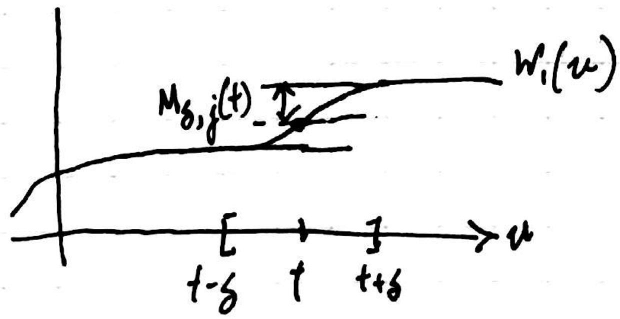
$\mu(t) = \mathbb{E}W_1(t)$.

Assume $\mathbb{E}\|W_1\|_\infty < \infty$.

Then $\bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow{p} \mu$ (here $W_i = l(\cdot; X_i) - l(\theta)$)

Proof. Fix $\varepsilon > 0$. NTS $P(\|W_n - \mu\|_\infty > \varepsilon) \to 0$.

Let $M_{\delta,j}(t) = \sup_{s:\|s-t\|<\delta} \left|W_j(s) - W_j(t)\right|$.



Define $\lambda_\delta(t) = \mathbb{E}[M_{\delta,1}(t)]$..

Lemma 9.1 says

$$\sup_t \lambda_\delta(t) \overset{\text{a.s.}}{\to} 0 \quad \text{as } \delta \downarrow 0 \tag{3.1}$$

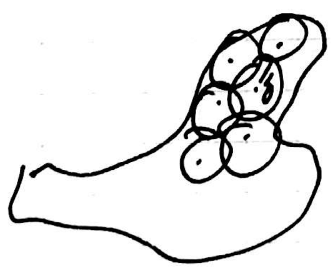
so we can choose $\delta > 0$ to ensure $\sup_{t\in K} \lambda_\delta(t) < \varepsilon$

Then if $\|s - t\| < \delta$, we have

$$\begin{aligned}
|\mu(t) - \mu(s)| &= |\mathbb{E}(W_1(t) - W_1(s))| && (3.2)\\
&\leq \mathbb{E}|W_1(t) - W_1(s)| && (3.3)\\
&\leq \mathbb{E}M_{\delta,j}(t) = \lambda_\delta(t) < \varepsilon && (3.4)
\end{aligned}$$

Let $B_\delta(t) = \{s : \|s - t\| < \delta\}$ (open ball).

Then $\{B_\delta(t) : t \in K\}$ is an open cover for K



K is compact, so can choose finite subcover $B_\delta(t_1), \ldots, B_\delta(t_m)$.

Let $\mathcal{O}_i = B_\delta(t_i) \cap K$.

$$\begin{aligned}
\|\bar{W}_n - \mu\|_\infty &= \max_{i=1}^{m} \sup_{s\in\mathcal{O}_i} |\bar{W}_n(s) - \mu(s)| && (3.5)\\
&\leq \max_i \sup_{s\in\mathcal{O}_i} \left[|\bar{W}_n(s) - \bar{W}_n(t_i)| + |\bar{W}_n(t_i) - \mu(t_i)| + \underbrace{|\mu(s) - \mu(t_i)|}_{<\varepsilon} \right] && (3.6)\\
&\leq \max_i \sup_{s\in\mathcal{O}_i} |\bar{W}_n(s) - \bar{W}_n(t_i)| + \max_i |\bar{W}_n(t_i) - \mu(t_i)| + \varepsilon && (3.7)
\end{aligned}$$

Now

$$\sup_{s\in\mathcal{O}_i}|\bar{W}_n(s)-\bar{W}_n(t_i)| = \frac{1}{n}\sup_{s\in\mathcal{O}_i}|\sum_{j=1}^{n} W_i(s)-W_j(t_i)| \tag{3.8}$$

$$\leq \frac{1}{n}\sum_{j=1}^{n}\sup_{s\in\mathcal{O}_i}|W_i(s)-W_i(t)| \tag{3.9}$$

$$= \frac{1}{n}\sum_{j=1}^{n} M_{\delta,j}(t_i) \xrightarrow{\text{p}} \lambda_\delta(t_i) < \varepsilon \tag{3.10}$$

Also notice

$$\|\bar{W}_n-\mu\|_\infty \leq 2\varepsilon + \max_i|\frac{1}{n}\sum M_{\delta,j}(t_i)-\lambda_s(t_i)| + \max_i|\bar{W}_n(t_i)-\mu(t_i)| \tag{3.11}$$

Hence

$$P(\|\bar{W}_n-\mu\|_\infty > 3\varepsilon) \to 0 \tag{3.12}$$

This shows the result for Θ compact. □

Theorem 3.2 (Keener 9.4). *Let $G_n \in C(K)$, $n \geq 1$ be random functions with $\|G_n - g\|_\infty \xrightarrow{p} 0$, some fixed $g \in C(K)$.*

Then

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

(b) If g is maximized at unique value t^* and $G_n(t_n) = \sup_{t\in K} G_n(t)$ then $t_n \xrightarrow{p} t^*$.

Proof. (a):

$$|G_n(t_n)-g(t^*)| \leq |G_n(t_n)-g(t_n)| + |g(t_n)-g(t^*)| \tag{3.13}$$

$$\leq \underbrace{\|G_n-g\|_\infty}_{\xrightarrow{\text{p}}0 \text{ by assumption}} + \underbrace{|g(t_n)-g(t^*)|}_{\substack{\xrightarrow{\text{p}}0 \\ \text{g cts}}} \tag{3.14}$$

□