Stat210A: Theoretical Statistics **Lecture Date: October 21st, 2014**

Random functions and MLE

Lecturer: Michael I. Jordan *Scribe: Richard Shin*

1 Random functions with random arguments

Last time, we were in the middle of talking about what happens if we have random functions with random arguments.

Theorem 1 (Theorem 9.4 of Keener (2010))**.** *$G_n \in C(K)$. Suppose that we have $\|G_n - g\|_\infty \xrightarrow{p} 0$ and $g \in C(k)$. Then*

- *If $t_n \xrightarrow{p} t^* \in K$, then $G_n(t_n) \xrightarrow{p} g(t^*)$.*
- *If g achieves its maximum at a unique value t^*, and if t_n maximizes G_n, then $t_n \xrightarrow{p} t^*$.*

Proof. For the first part:

$$\begin{aligned}|G_n(t_n) - g(t^*)| &\le |G_n(t_n) - g(t_n)| + |g(t_n) - g(t^*)| \quad \text{(triangle inequality)}\\ &\le \|G_n - g\|_\infty + |g(t_n) - g(t^*)|\end{aligned}$$

We also know that $g(t_n) \xrightarrow{p} g(t^*)$. Then

$$\begin{aligned}\Rightarrow P(|G_n(t_n) - g(t^*)| > \epsilon) &\le P(\|G_n - g\|_\infty + |g(t_n) - g(t^*)| > \epsilon)\\ &\le P(\underbrace{\|G_n - g\|_\infty}_{Z_1} > \frac{\epsilon}{2}) + P(\underbrace{|g(t_n) - g(t^*)|}_{Z_2} > \frac{\epsilon}{2})\end{aligned}$$

From assumptions we have that $\|G_n - g\|_\infty \xrightarrow{p} 0$ and $g(t_n) \xrightarrow{p} g(t^*)$, so we are done.

We used the union bound to break up the probability. Recall the the union bound is $P(A \cup B) \le P(A) + P(B)$.

$$P(Z_1 + Z_2) > \epsilon$$
$$\{Z_1 + Z_2 > \epsilon) \Rightarrow \{Z_1 > \frac{\epsilon}{2}\} \cup \{Z_2 > \frac{\epsilon}{2}\}$$

For the second part:
Fix $\epsilon > 0$. Let $K_\epsilon = K - B_\epsilon(t^*)$, and

$$\begin{gathered}M = g(t^*)\\ M_\epsilon = \sup_{t \in K_\epsilon} g(t)\\ K_\epsilon \text{ compact} \Rightarrow M_\epsilon = g(t^*_\epsilon) \quad t^*_\epsilon \in K_\epsilon\\ \text{and } M_\epsilon < M\end{gathered}$$

Let $\delta = M - M_\epsilon$, and suppose $\|G_n - g\|_\infty < \frac{\delta}{2}$.

$$
\begin{aligned}
(*) \Rightarrow \sup_{K_\epsilon} G_n &< \sup_{K_\epsilon} g + \frac{\delta}{2} = M_\epsilon + \frac{\delta}{2} = M - \frac{\delta}{2} \\
\Rightarrow \sup_{K} G_n &\geq G_n(t^*) > g(t^*) - \frac{\delta}{2} = M - \frac{\delta}{2} \\
\Rightarrow \sup_{K} G_n &\geq M - \frac{\delta}{2} > \sup_{K_\epsilon} G_n \\
&\Rightarrow t_n \text{, which maximizes } G_n \text{, lies in } B_\epsilon(t^*) \\
&\Rightarrow P(\|G_n - g\|_\infty < \frac{\delta}{2}) \leq P(\|t_n - t^*\| < \epsilon) \\
&\Rightarrow P(\|t_n - t^*\| \geq \epsilon) \leq P(\|G_n - g\|_\infty \geq \frac{\delta}{2}) \to 0
\end{aligned}
$$

□

2 Consistency of MLE

Assume that $X, X_1, X_2, \cdots$ are i.i.d. from f_θ (continuous in θ).

$$
l_n(\omega) = \log \prod_{i=1}^{n} f_\omega(X_i) = \sum_i \log f_\omega(X_i)
$$

$$
\hat{\theta}_n \in \arg\max\, l_n(\omega)
$$

The Kullback-Leibler divergence is

$$
I(\theta, \omega) = E_\theta \log \frac{f_\theta(X)}{f_\omega(X)}
$$

$$
I(\theta, \omega) > 0 \qquad \text{unless } \theta = \omega
$$

Let us also define

$$
W(\omega) = \log \frac{f_\omega(X)}{f_\theta(X)}
$$

Theorem 2 (Theorem 9.9 of Keener (2010))**.** *Ω compact, $E_\theta \|\omega\|_\infty < \infty$, $f_\omega(x)$ is continuous in w a.e. x, and $P_\omega \neq P_\theta$ if $\theta \neq \omega$ (identifiability). Then*

$$
\hat{\theta}_n \xrightarrow{p} \theta.
$$

Proof. Let $W_i(\omega) = \log \frac{f_\omega(X_i)}{f_\theta(X_i)} \in C(\Omega)$. $W_i(\omega)$ are i.i.d. with mean $-I(\theta, \omega) = \mu(\omega)$. This has a unique maximum at θ.

Let $\bar{W}_n(\omega) = \frac{1}{n} \sum_i W_i(\omega) = \frac{1}{n} l_n(\omega) - \frac{1}{n} l_n(\theta)$. $\hat{\theta}_n$ maximizes $\bar{\omega}_n(\omega)$.

Theorem 9.2 implies $\|\bar{W}_n - \mu\|_\infty \xrightarrow{p} 0$ and Theorem 9.4(1) implies $\hat{\theta}_n \xrightarrow{p} \theta$. □

Theorem 3 (Theorem 9.9, without compactness)**.** *Let $\Omega = \mathbf{R}^p$, let $f_\omega(x)$ be continuous in ω a.e. x. Let $P_\theta \neq P_\omega$ for $\theta \neq \omega$, let $f_\omega(x) \to 0$ as $\omega \to \infty$ a.e. x. If $E_\theta \|\mathbf{1}_K W\|_\infty < \infty$ for all compact $K \subseteq \mathbf{R}^p$, and if $E_\theta \sup_{\|\omega\| > a} W(\omega) < \infty$ for some a, then*

$$
\hat{\theta}_n \xrightarrow{p} \theta.
$$

See Keener (2010) for the proof.

3 Distributional results

Lemma 4 (Lemma 9.15 of Keener (2010)). *Suppose $Y_n \Rightarrow Y$ and $P(B_n) \to 1$. Then, for arbitrary RVs $_n$,*

$$Y_n \mathbf{1}_{B_n} + Z_n \mathbf{1}_{B_n^C} \Rightarrow Y.$$

Proof. Let $\epsilon > 0$.

$$\begin{aligned} P(|Z_n \mathbf{1}_{B_n^C}| > \epsilon) &\le P(B_n^C) = 1 - P(B_n) \to 0 \\ P(|\mathbf{1}_{B_n} - \mathbf{1}| > \epsilon) &\le P(B_n^C) \to 0 \\ &\Rightarrow \mathbf{1}_{B_n} \xrightarrow{p} 1 \end{aligned}$$

Using Slutsky, $Y_n \mathbf{1}_{B_n} + Z_n \mathbf{1}_{B_n^C} \Rightarrow Y$. □

We now define the following notation:

- $W(\theta) = \log f_\theta(X)$
- $I(\theta) = E_\theta (W'(\theta))^2 = -E_\theta W''(\theta)$
- $E_\theta W'(\theta) = 0$

Remark 5 (Statement 5 of Theorem 9.14). $\forall \theta \in \Omega^0, \exists \epsilon > 0$ s.t. $E_\theta \| \mathbf{1}_{(\theta-\epsilon,\theta+\epsilon)} W'' \|_\infty < \infty$. Then

$$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N\left(0, \frac{1}{I(\theta)}\right) \quad \theta \in \Omega^0$$

Proof. Use this statement to choose $\epsilon > 0$ s.t. $E_\theta \| \mathbf{1}_{(\theta-\epsilon,\theta+\epsilon)} W'' \|_\infty < \infty$ and $[\theta - \epsilon, \theta + \epsilon] \subset \Omega^0$. Let B_n denote the event that $\hat{\theta}_n \in (\theta - \epsilon, \theta + \epsilon)$.

$$\text{Consistency} \Rightarrow P(B_n) \to 1.$$

Define $\bar{W}_n(\omega) = \frac{1}{n} l_n(\omega) = \frac{1}{n} \sup_i \log f_\omega(X_i)$. Taking the Taylor expansion of $\bar{W}_n'$,

$$\begin{aligned} \bar{W}_n'(\hat{\theta}_n) &= \bar{W}_n'(\theta) + \bar{W}_n''(\tilde{\theta}_n)(\hat{\theta}_n - \theta) = 0 \\ \sqrt{n}(\hat{\theta}_n - \theta) &= \frac{\sqrt{n}\bar{W}_n'(\theta)}{-\bar{W}_n''(\tilde{\theta}_n)} \\ \text{CLT} \Rightarrow \sqrt{n}\bar{W}_n'(\theta) &\Rightarrow N(0, I(\theta)) \end{aligned}$$

If the denominator converges in probability to $I(\theta)$, we're done (Slutsky), since if $Y = aX$ then $\text{Var}Y = a^2 \text{Var}X$.

$$\begin{aligned} \text{On } B_n \quad |\tilde{\theta}_n - \theta| \le |\hat{\theta}_n - \theta| &\Rightarrow \tilde{\theta}_n \xrightarrow{p} \theta \\ \text{Theorem 9.2} &\Rightarrow \| \mathbf{1}_{(\theta-\epsilon,\theta+\epsilon)} (\bar{W}_n'' - \mu) \|_\infty \xrightarrow{p} 0 \\ \mu(\omega) &= E_\theta W''(\omega) \\ \text{Theorem 9.4 part (1)} &\Rightarrow \bar{W}_n''(\tilde{\theta}_n) \to \mu(\theta) = -I(\theta) \end{aligned}$$

□

This was a taste of the harder parts of empirical process theory.

References

Keener, R. (2010). *Theoretical Statistics: Topics for a Core Course*. Springer, New York, NY.