Stats 300b: Theory of Statistics **Winter 2017**

Lecture 16 – March 2, 2017

Lecturer: John Duchi *Scribe: Jaime Roquero Gimenez*

Warning: *these notes may contain factual errors*

Reading:

Outline:

1. Finish uniform laws, rates and moduli of continuity.

2. Asymptotic testing: Power, Local alternatives, Efficiency and comparing tests

1 Recap: M-estimation problem

We consider the objective function and its empirical form

$$\begin{aligned} M_n &: \Theta \to \mathbb{R} \\ M &: \theta \mapsto \mathbb{E}[M_n(\theta)] \end{aligned}$$

and we consider the case where $M(\theta_0) \geq M(\theta) + d(\theta, \theta_0)^2$ near $\theta_0 = \operatorname{argmax}_{\theta \in \Theta} M(\theta)$.

Idea: compare $M(\theta_0) - M(\theta)$ (Gaps in population) vs. $M_n(\theta_0) - M_n(\theta)$ (Gaps in empirical version).

Example 1: Suppose we had

$$\sup_{d(\theta_0,\theta) \leq \delta} |M_n(\theta) - M(\theta) - (M_n(\theta_0) - M(\theta_0))| \leq \frac{\sigma\delta}{\sqrt{n}}$$

Let $\hat{\theta}_n = \operatorname{argmax}_\theta M_n(\theta)$, assume consistency of the estimator: ie. $\hat{\theta}_n \to \theta_0$ in probability. Then

$$\begin{aligned} M_n(\hat{\theta}_n) \geq M_n(\theta_0) &= M_n(\theta_0) - M(\theta_0) + M(\hat{\theta}_n) + M(\theta_0) - M(\hat{\theta}_n) \\ &\geq M_n(\theta_0) - M(\theta_0) + M(\hat{\theta}_n) + d(\theta_0, \hat{\theta}_n)^2 \end{aligned}$$

$$\begin{aligned} 0 &\geq M_n(\theta_0) - M(\theta_0) - (M_n(\hat{\theta}_n) - M(\hat{\theta}_n)) \\ 0 &\geq -\frac{\sigma d(\theta_0, \hat{\theta}_n)}{\sqrt{n}} + d(\theta_0, \hat{\theta}_n)^2 \end{aligned}$$

ie. $d(\theta_0, \hat{\theta}_n) \leq \frac{\sigma}{\sqrt{n}}$
♣

Example 2: Let $m_\theta(x)$ be Lipschitz in θ, and $M_n(\theta) = P_n m_\theta(x)$.

Let $P_n^0 = \frac{1}{n}\sum_{i=1}^n \epsilon_i \delta_{X_i}$ the symmetrized empirical measure.

Then

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{d(\theta,\theta_0)\leq\delta} \big|M_n(\theta) - M(\theta) - (M_n(\theta_0) - M(\theta_0))\big|\Big] \\
(\star) &\leq 2\,\mathbb{E}\,[\sup_{d(\theta,\theta_0)\leq\delta} P_n^0(m_\theta - m_{\theta_0})(X)] \\
(\star\star) &\leq \frac{Cst}{\sqrt{n}}\int_0^\delta \sqrt{\log N(\delta - ball, d, \epsilon)}\, d\epsilon \\
&\leq \frac{Cst}{\sqrt{n}}\int_0^\delta \sqrt{d\log(1 + \frac{2\delta}{\epsilon})}\, d\epsilon \\
&\leq Cst\,\frac{\delta\sqrt{d}}{\sqrt{n}}
\end{aligned}
$$

Where in $(\star)$ we apply symmetrization and in $(\star\star)$ we bound by the entropy integral given that

$$
\mathbb{E}\left[\exp\Big(\lambda\sum_{i=1}^n \epsilon_i\big(m_\theta(X_i) - m_{\theta_0}(X_i)\big)\Big)\right] \leq \exp\Big(\frac{\lambda^2}{2} n d(\theta,\theta_0)^2\Big)
$$

as $m_\theta(x)$ is Lipschitz in θ, so $\sqrt{n}P_n^0(m_\theta - m_{\theta_0})$ is d^2 sub-gaussian.

♣

2 Moduli of continuity and convergence rates

Theorem 1. *Suppose $M(\theta_0) \geq M(\theta) + d(\theta,\theta_0)^2$ near θ_0.*

Let ϕ be such that $\phi(c\delta) \leq c^\alpha\phi(\delta)$ for some $\alpha \in (0,2)$.

Assume

$$
\mathbb{E}\Big[\sup_{d(\theta,\theta_0)\leq\delta} \big|M_n(\theta) - M(\theta) - (M_n(\theta_0) - M(\theta_0))\big|\Big] \leq \frac{\phi(\delta)}{\sqrt{n}}
$$

Let $r_n \to +\infty$ such that $r_n^2\phi(\frac{1}{r_n}) \leq \sqrt{n}$.

If $\hat\theta_n \to \theta_0$ in probability, then $r_n d(\hat\theta_n,\theta_0) = O_P(1)$

Idea: Estimation errors can scale at most as $\frac{\phi(\delta)}{\sqrt{n}} \sim \frac{d(\theta,\theta_0)^\alpha}{\sqrt{n}}$ with $\alpha < 2$. But the growth of the objective function is quadratic: $d(\theta,\theta_0)^2$.

We solve $\delta^2 = \frac{\delta^\alpha}{\sqrt{n}}$: it implies $\delta = \big(\frac{1}{n}\big)^{\frac{1}{2(2-\alpha)}}$

Proof Let $\eta \leq \delta$. Then $\mathbb{P}\big(d(\hat\theta_n,\theta_0) \geq \eta\big) = o(1)$ (consistency of the estimator).

Define the shell: $S_{nj} = \{\theta : 2^{j-1} \leq r_n d(\theta,\theta_0) \leq 2^j\}$.

Consider the event

$$
d(\hat\theta_n,\theta_0) \geq \frac{2^t}{r_n}
$$

for some $t \geq 0$.

Then $\exists j \geq t$ such that $\hat{\theta}_n \in S_{jn}$.
Therefore

$$\begin{aligned}
\mathbb{P}(r_n d(\hat{\theta}_n, \theta_0) \geq 2^t) &\leq \mathbb{P}(r_n d(\hat{\theta}_n, \theta_0) \geq 2^t, d(\hat{\theta}_n, \theta_0) \leq \eta) + \mathbb{P}(d(\hat{\theta}_n, \theta_0) > \eta) \\
&= \Big(\sum_{j \leq t,\, 2^{j-1} \leq r_n \eta} \mathbb{P}(\hat{\theta}_n \in S_{jn}) \Big) + \mathbb{P}(d(\hat{\theta}_n, \theta_0) > \eta) \\
&\leq \Big(\sum_{j \leq t,\, 2^{j-1} \leq r_n \eta} \mathbb{P}(\exists \theta \in S_{jn}, M_n(\theta) \geq M_n(\theta_0)) \Big) + o(1)
\end{aligned}$$

But if $M_n(\theta) \geq M_n(\theta_0)$ for some $\theta \in S_{jn}$ then

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

So

$$\sup_{\theta \in S_{jn}} \big| (M_n(\theta) - M(\theta)) - (M_n(\theta_0) - M(\theta_0)) \big| \geq \frac{2^{2j}}{r_n^2}$$

And

$$\begin{aligned}
\mathbb{P}(\exists \theta \in S_{jn}, M_n(\theta) \geq& M_n(\theta_0)) \\
&\leq \mathbb{P}\Big(\sup_{\theta \in S_{jn}} \big| (M_n(\theta) - M(\theta)) - (M_n(\theta_0) - M(\theta_0)) \big| \geq \frac{4^j}{r_n^2} \Big) \\
&\leq \frac{r_n^2}{4^j} \mathbb{E}\Big[\sup_{\theta \in S_{jn}} \big| (M_n(\theta) - M(\theta)) - (M_n(\theta_0) - M(\theta_0)) \big| \Big] \\
&\leq \frac{r_n^2}{4^j \sqrt{n}} \phi\big(\frac{2^j}{r_n}\big) \;\leq\; \frac{r_n^2\, 2^{\alpha j}}{4^j \sqrt{n}} \phi\big(\frac{1}{r_n}\big) \;\leq\; 2^{j(\alpha - 2)}
\end{aligned}$$

So we bound the inequality above with this result and get:

$$\mathbb{P}(d(\hat{\theta}_n, \theta_0) \geq \frac{2^t}{r_n}) \leq \sum_{j \geq t} 2^{-j(2-\alpha)} + o(1)$$

So by taking t large enough we get that $r_n d(\hat{\theta}_n, \theta_0) = O_P(1)$

□

Observation 2. *If $\phi(\delta) = \delta^\alpha$, solving $r_n^2 \phi(\frac{1}{r_n}) = \sqrt{n}$ gives $r_n = n^{\frac{1}{2(2-\alpha)}}$ for the convergence rate.*

More generally, if $M(\theta_0) \geq M(\theta) + d(\theta, \theta_0)^\beta$ then we could choose r_n such that $r_n^\beta \phi(\frac{1}{r_n}) \leq \sqrt{n}$ so that $r_n = n^{\frac{1}{2(\beta-\alpha)}}$. In order to get $r_n \to +\infty$, we want cases where growth in $d(.,.)^\beta$ dominates moduli of continuity $\phi(\delta) \sim \delta^\alpha$ at size $\sqrt{n}$.