Lecture 4: February 28

Lecturer: Anirban Bhattacharya & Debdeep Pati *Scribes: Indrajit Ghosh & Xiaomeng Yan*

Note: *LaTeX template courtesy of UC Berkeley EECS dept & CMU's convex optimization course taught by Ryan Tibshirani.*

4.1 Proof of (an extension of) Schwartz's Theorem

Assume p_0 is the true density and $p_0 \in \mathrm{KL}(\pi)$. There exists a test function $\mathbf{\Phi}_n$ such that

$$\mathbb{E}_{p_0}\mathbf{\Phi}_n \le e^{-cn}$$
$$\sup_{p\in\mathcal{U}^c\cap\mathcal{P}_n} \mathbb{E}_p(1-\mathbf{\Phi}_n) \le e^{-cn}$$

for some constant $c > 0$, where $\mathcal{P}_n$ is a measurable subset of $\mathcal{P}$ such that $\pi(\mathcal{P}_n^c) < e^{-cn}$.

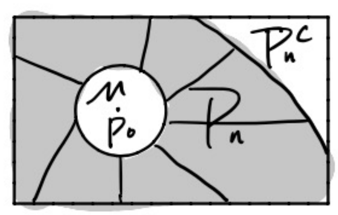

Figure 4.1:

The sets $\{\mathcal{P}_n\}$ are called "Sieves" and $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ for large n.

Lemma 4.1. *(Borel-Centelli lemma) If V_n is a sequence of random variabless such that for any $\epsilon > 0$,*

$$\sum_{n=1}^{\infty} P(|V_n| > \epsilon) < \infty,$$

then $V_n \to 0$ a.s.

Proof. **Goal**: show that $\pi_n(\mathcal{U}^c|X^{(n)}) \to 0$ a.s. $[p_0]$. For any subset of densities $\mathcal{B}$ (that is $\mathcal{B} \subset \mathcal{P}$), the posterior is defined as

$$\pi_n(\mathcal{B}|X^{(n)}) = \frac{\int_{\mathcal{B}} \prod_{i=1}^n p(X_i)\pi(dp)}{\int_{\mathcal{P}} \prod_{i=1}^n p(X_i)\pi(dp)} = \frac{\int_{\mathcal{B}} \prod_{i=1}^n \frac{p(X_i)}{p_0(X_i)}\pi(dp)}{\int_{\mathcal{P}} \prod_{i=1}^n \frac{p(X_i)}{p_0(X_i)}\pi(dp)}$$

Note that

$$\begin{aligned}\pi_n(\mathcal{U}^c|X^{(n)}) =& \pi_n(\mathcal{U}^c \cap \mathcal{P}_n|X^{(n)}) + \pi_n(\mathcal{U}^c \cap \mathcal{P}_n^c|X^{(n)}) \\ \leq& \pi_n(\mathcal{U}^c \cap \mathcal{P}_n|X^{(n)}) + \pi_n(\mathcal{P}_n^c|X^{(n)}) \\ =& \mathbf{\Phi}_n \pi_n(\mathcal{U}^c \cap \mathcal{P}_n|X^{(n)}) + (1-\mathbf{\Phi}_n)\pi_n(\mathcal{U}^c \cap \mathcal{P}_n|X^{(n)}) + \pi_n(\mathcal{P}_n^c|X^{(n)}) \\ \leq& \mathbf{\Phi}_n + (1-\mathbf{\Phi}_n)\pi_n(\mathcal{U}^c \cap \mathcal{P}_n|X^{(n)}) + \pi_n(\mathcal{P}_n^c|X^{(n)})\end{aligned}$$

Claim 1: $\mathbf{\Phi}_n \to 0$ a.e $[p_0]$.

Proof. Fix $\delta > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}_0[\mathbf{\Phi}_n > \delta] \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}_{p_0[\mathbf{\Phi}_n]}}{\delta} \leq \frac{1}{\delta}\sum_{n=1}^{\infty} e^{-cn} < \infty.$$

The first inequality is given by Markov inequality. Then according to Borel-Cantelli lemma,

$$\mathbf{\Phi}_n \to 0 \text{ a.s.}$$

□

Claim 2: For any $\epsilon > 0$, define $\mathcal{P}_\epsilon = \{p \in \mathcal{P}, D(p_0\|p) < \epsilon\}$. Then

$$\int_{\mathcal{P}} \prod_{i=1}^{n} \frac{p(X_i)}{p_0(X_i)}\pi(dp) \geq \pi(\mathcal{P}_\epsilon)e^{-n\epsilon} \text{ eventually a.e } [p_0]$$

Proof.

$$\int_{\mathcal{P}} \prod_{i=1}^{n} \frac{p(X_i)}{p_0(X_i)}\pi(dp) \geq \pi(\mathcal{P}_\epsilon)\int_{\mathcal{P}_\epsilon} \prod_{i=1}^{n} \frac{p(X_i)}{p_0(X_i)}\pi_\epsilon(dp)$$

where π_ϵ is the "truncation" of π on $\mathcal{P}_\epsilon$, that is $\pi_\epsilon(\mathcal{B}) = \frac{\pi(\mathcal{B}\cap\mathcal{P}_\epsilon)}{\pi(\mathcal{P}_\epsilon)}$.

Consider

$$\log(\int_{\mathcal{P}_\epsilon} \prod_{i=1}^{n} \frac{p(X_i)}{p_0(X_i)}\pi_\epsilon(dp)) \geq n\int_{\mathcal{P}_\epsilon} \frac{1}{n}\sum_{i=1}^{n} \log\frac{p(X_i)}{p_0(X_i)}\pi_\epsilon(dp)$$

by Jensen's inequality.

Now

$$\int_{\mathcal{P}_\epsilon} \frac{1}{n}\sum_{i=1}^{n} \log\frac{p(X_i)}{p_0(X_i)}\pi_\epsilon(dp) \xrightarrow{\text{a.s.}} \int_{\mathcal{P}_\epsilon} D(p_0\|p)\pi_\epsilon(dp) \geq -\epsilon$$

by construction of $\mathcal{P}_\epsilon$. This proves claim 2.

□

Now back to proof of (an extension of) Schwartz's Theorem.

$$(1-\mathbf{\Phi}_n)\pi_n(\mathcal{U}^c\cap\mathcal{P}_n|X^{(n)}) \leq (1-\mathbf{\Phi}_n)\frac{e^{n\epsilon}}{\pi(\mathcal{P}_\epsilon)}\int_{\mathcal{U}^c\cap\mathcal{P}_n}\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\pi(dp)$$
$$= ce^{n\epsilon}(1-\mathbf{\Phi}_n)\int_{\mathcal{U}^c\cap\mathcal{P}_n}\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\pi(dp)$$

Consider

$$\begin{aligned}
&\mathbb{E}_{p_0}[(1-\mathbf{\Phi}_n)\int_{\mathcal{U}^c\cap\mathcal{P}_n}\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\pi(dp)]\\
=&\int_{\mathcal{U}^c\cap\mathcal{P}_n}[\int(1-\mathbf{\Phi}_n)\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\prod_{i=1}^n p_0(X_i)dX_i]\pi(dp) \text{ Fubini's theorem}\\
=&\int_{\mathcal{U}^c\cap\mathcal{P}_n}[\int(1-\mathbf{\Phi}_n)\prod_{i=1}^n p(X_i)dX_i]\pi(dp)\\
=&\int_{\mathcal{U}^c\cap\mathcal{P}_n}\mathbb{E}_p(1-\mathbf{\Phi}_n)\pi(dp)\\
\leq& e^{-nc}
\end{aligned}$$

This implies $(1-\mathbf{\Phi}_n)\pi_n(\mathcal{U}^c\cap\mathcal{P}_n|X^{(n)})\to 0$ [a.e.] if we take $\epsilon=\frac{c}{2}$

$$\pi_n(\mathcal{P}_n^c|X^{(n)}) \leq ce^{n\epsilon}\int_{\mathcal{P}_n^c}\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\pi_\epsilon(dp)$$

Consider

$$\begin{aligned}
&\mathbb{E}_{p_0}[\int_{\mathcal{P}_n^c}\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\pi(dp)]\\
=&\int_{\mathcal{P}_n^c}[\int\prod_{i=1}^n\frac{p(X_i)}{p_0(X_i)}\prod_{i=1}^n p_0(X_i)]\pi(dp) \text{ Fubini's theorem}\\
=&\int_{\mathcal{P}_n^c}\prod_{i=1}^n p(X_i)\pi(dp)\\
=&\pi(\mathcal{P}_n^c)\leq e^{-nc}
\end{aligned}$$

This implies $\pi_n(\mathcal{P}_n^c|X^{(n)})\to 0$ [a.e.] if we take $\epsilon=\frac{c}{2}$.

□

Question: How to construct such test?

$$H_0: p = p_0$$
$$H_1: p\in\cup_{i=1}^N\mathcal{B}_\epsilon(p_i)$$

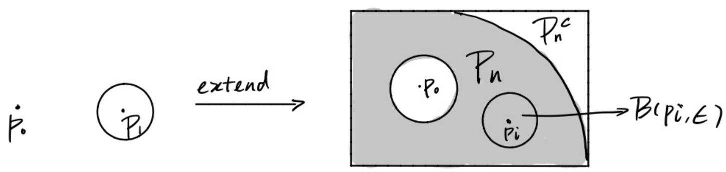

Figure 4.2:

where $\mathcal{B}_\epsilon(p_i) = \{p : \|p - p_i\|_{\mathrm{TV}} < \frac{\|p-p_0\|_{\mathrm{TV}}}{2}\}$.

Let $\mathbf{\Phi}_i$ be an usual test function for p_0 against $\mathcal{B}_\epsilon(p_i)$. Define $\mathbf{\Phi} = \max\{\mathbf{\Phi}_1, ..., \mathbf{\Phi}_N\}$ (that is taking the union of the rejection region), we need to control over N: number of balls to cover $\mathcal{U}^c \cap \mathcal{P}_n$.