Lecture 3: February 26

Lecturer: Anirban Bhattacharya & Debdeep Pati *Scribes: Eric Chuu, Pallavi Ray*

3.1 Example (Discrete Paramter Space)

Let $\Theta = \{\theta_1, \ldots, \theta_K\}, X_1, \ldots, X_n \overset{\text{iid}}{\sim} p_{\theta_j}$, where θ_j is the true parameter. Denote $p_{\theta_j}(x) = p(x \mid \theta_j)$ and let $\pi_j = \pi(\theta = \theta_j)$ be the prior, for $j = 1, \ldots K$. Here, we also denote $X^{(n)} = (X_1, \ldots, X_n)$. We want to show that for $a \neq j$, $\pi\left(\theta_a \mid X^{(n)}\right) \overset{\text{a.s.}}{\to} 0$, as this would imply that $\pi\left(\theta_a \mid X^{(n)}\right) \overset{\text{a.s.}}{\to} 0$ under p_{θ_j}. First suppose $\pi_j = 0$. Then $\pi\left(\theta_j \mid X^{(n)}\right) \equiv 0$. For posterior consistency, we assume that $\pi_j > 0$ for all $j = 1, \ldots, K$. Then, for $a \neq j$, consider

$$\begin{aligned}
\log \frac{\pi\left(\theta_a \mid X^{(n)}\right)}{\pi\left(\theta_j \mid X^{(n)}\right)} &= \log \frac{\pi_a}{\pi_j} + \log \sum_{i=1}^n \frac{p\left(X_i \mid \theta_a\right)}{p\left(X_i \mid \theta_j\right)} \\
&= n\left(\frac{1}{n} \log \frac{\pi_a}{\pi_j} + \frac{1}{n} \log \sum_{i=1}^n \frac{p\left(X_i \mid \theta_a\right)}{p(X_i \mid \theta_j)}\right)
\end{aligned}$$

Note that under p_{θ_j},

$$\frac{1}{n} \log \frac{\pi_a}{\pi_j} \to 0$$

By SLLN,

$$\frac{1}{n} \log \sum_{i=1}^n \frac{p\left(X_i \mid \theta_a\right)}{p(X_i \mid \theta_j)} \overset{\text{a.s.}}{\to} -D(p_{\theta_j} \parallel p_{\theta_a})$$

Using these two observations, we have the following

$$\log \frac{\pi\left(\theta_a \mid X^{(n)}\right)}{\pi\left(\theta_j \mid X^{(n)}\right)} \overset{\text{a.s.}}{\to} -\infty \quad [p_{\theta_j}]$$

and we conclude $\pi\left(\theta_j \mid X^{(n)}\right) \overset{\text{a.s.}}{\to} 1, \left[p_{\theta_j}\right]$.

Remark

1. It is assumed that p_θ does not depend on n.
2. If $\pi_j = 0$, then there is no posterior consistency.
3. Case of non-identifiability: $\theta_a \neq \theta_j$ but $D(p_{\theta_j} \parallel p_{\theta_a}) = 0$. THen there is no posterior consistency.
4. The complexity of the parameter space should be sufficiently small.

3.2 Some important theorems and concepts

Theorem 3.1. *(Doob) Let* $(\Omega, \mathcal{A}, p_\theta), \theta \in \Theta$. *Assume* $\theta \mapsto p_\theta$ *is 1-1. Let* $X_1 \dots X_n \overset{iid}{\sim} p_\theta, \theta \sim \pi$. *Then* $\pi_n\left(\cdot \mid X^{(n)}\right)$ *is strongly consistent for* π*-almost* θ. *That is, for* $\Theta_0 \subseteq \Theta$ *such that* $\pi(\Theta_0) = 1$, *then* $\pi_n\left(\cdot \mid X^{(n)}\right)$ *is strongly consistent if* $X_1, \dots, X_n \overset{iid}{\sim} p_{\theta_0}, \theta_0 \in \Theta_0$.

Remark. If $\pi(\{\theta_0\}) = 0$, or π depends on n, then the theorem says nothing. The theorem applies for a countable parameter space, provided that $\theta \mapsto p_\theta$ is 1-1.

Definition 3.2. *(Kullback–Leibler property and Kullback–Leibler support)*
Let $\mathcal{P}$ *be density space and* $p_0 \in \mathcal{P}$. p_0 *is said to have KL property relative to a prior* π *if*

$$\pi\{p : D(p_0 \parallel p) < \epsilon\} > 0 \text{ for every } \epsilon > 0$$

And KL support, $KL(\pi)$ *, is given by :*

$$KL(\pi) = \{p_0 \in \mathcal{P} : p_0 \text{ has KL property}\}$$

Theorem 3.3. *(Schwartz) If* $p_0 \in KL(\pi)$ *and* $\mathcal{U}$ *is a neighborhood around* p_0 *such that there exists a test function* $\phi_n = \phi_n(X_1, \dots, X_n)$, *and* $\mathbb{E}_{p_0}\phi_n \to 0$ *and* $\sup_{p \in \mathcal{U}^c} \mathbb{E}_p(1 - \phi_n) \to 0$ *as* $n \to \infty$, *then*
$\pi_n\left(\mathcal{U}^c \mid X^{(n)}\right) \to 0$, *a.s.* $[p_0]$

- A typical choice of $\mathcal{U} = \{p : d(p, p_0) < \epsilon\}$
- If $\mathcal{U}$ is weak–neighbourhood, then ϕ_n satisfying the above, always exists.

Definition 3.4. $P_n \overset{w}{\Longrightarrow} P$ *iff for any bounded continuous function* ϕ, $\int \phi dP_n \to \int \phi dP$
$\iff P_n(-\infty, x] \to P(-\infty, x]$, $\forall\, x \in$ *continuity points of* $P(-\infty, x]$
$\iff P_n(A) \to P(A)\ \forall\ A$ *s.t.* $P(\partial A) = 0$

Consider the metric $(\mathbb{R}, d \mid x - y \mid)$. Then $\{(-\infty, x] : x \in \mathbb{R}\}$ is called "generating set" or " building block" for open sets in $\mathcal{B}(\mathbb{R})$. The following result characterize an open set around p_0.

Result : A generating set of a weak–neighbourhood around p_0 is given by

$$\left\{\left\{P : \int \phi\, dP < \int \phi\, dP_0 + C\right\}, \phi \text{ is bounded and continuous and } C > 0\right\}$$

Definition 3.5. *(Levy metric) Let* F *and* G *be distribution functions.*

$$d(F, G) = \inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \le G(x) \le F(x + \epsilon) + \epsilon, \forall x \in \mathbb{R}\}$$

Let us fix ϕ (bounded and continuous) and $C > 0$ and define

$$\mathcal{U} = \left\{P : \int \phi\, dP < \int \phi\, dP_0 + C\right\}$$

The goal is to construct ϕ_n based on $\{X_1, \dots, X_n\}$ such that

$$\mathbb{E}_{p_0}\phi_n \to 0 \text{ and } \sup_{p \in \mathcal{U}^c} \mathbb{E}_p(1 - \phi_n) \to 0 \text{ as } n \to \infty \tag{3.1}$$

Let us start with X_1 and we shall construct an unbiased test based on X_1.

Without loss of generality, assume $0 \leq \phi \leq 1$.

In $\mathcal{U}^c$: $\int \phi(X)\, dP(X) < \int \phi(X)\, dP_0(X) + C \implies \phi$ is an unbiased test.

Now, let us define ϕ_n based $\{X_1, \ldots, X_n\}$ as: $\phi_n = \mathbb{1}\left\{\frac{1}{n}\sum_{i=1}^{n} \phi(X_i) > \mathbb{E}_{p_0}\phi + \frac{C}{2}\right\}$

It is to be noted that ϕ_n such constructed satisfies (3.1).