Lecture 6

Agenda: * Scheffé Selector/Estimator

[A] Scheffé Selector

Problem: Given samples $X_1, \dots, X_n$ from P, find i s.t.

$$d(P_i, P) \le c \cdot \min_j d(P_j, P).$$

Let $P_1, \dots, P_M$ and P have densities $f_1, \dots, f_M$ and f w.r.t. ν, i.e.,

$$P_i(A) = \int_A f_i(x)\, \nu(dx)$$

Note that

$$d(P_i, P_j) = \frac{1}{2} \int |f_i(x) - f_j(x)|\, \nu(dx).$$

Consider $M=2$ first.

Denote by μ_n the empirical measure

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \in A\}.$$

The Scheffe selector is given by

$$\hat{f}(X^n) = \operatorname*{argmin}_{i=1,2} \left| \int_{A^*} f_i(x)\nu(dx) - \mu_n(A^*) \right|,$$

where

$$A^* \equiv A^*(f_1, f_2) = \{x : f_1(x) \ge f_2(x)\}.$$

Theorem

$$d(\hat{P}, P) \le 3 \min\{d(P_1, P), d(P_2, P)\} + 2\left|P^n(A^*) - \mu_n(A^*)\right|.$$

Proof. Consider $\bar{P} = \underset{i=1,2}{\operatorname{argmin}}\ d(P_i, P)$ and denote by $\bar{f}$ the density of $\bar{P}$.

$$d(\hat{P}, P) \le d(\hat{P}, \bar{P}) + d(P, \bar{P})$$

$$d(\hat{P}, \bar{P}) = \mathbb{1}_{\{\hat{P}=P_1,\, \bar{P}=P_2\}}\, d(P_1, P_2) + \mathbb{1}_{\{\hat{P}=P_2,\, \bar{P}=P_1\}}\, d(P_1, P_2)$$

Consider the first term. It evaluates the case $\bar{f} = f_2$. Note that since $\{f_1, f_2\}$ are allowed to depend on X^n, the event $\{\bar{f} = f_2\}$ is random. Denote by A the set $\{x : f_1(x) \ge f_2(x)\}$ and $E = \{\hat{f} = f_1, \bar{f} = f_2\}$. Then,

$$\mathbb{1}_E\, d(P_1, P_2) \le \mathbb{1}_E\left[(P_1(A) - \mu_n(A)) - (P_2(A) - \mu_n(A))\right]$$

Now consider the following cases: (note $P_1(A) \ge P_2(A)$)

(i) $\underline{P_1(A) \ge \mu_n(A) \ge P_2(A)}$

$$|P_1(A) - \mu_n(A)| = P_1(A) - \mu_n(A)$$

$(E \Rightarrow \hat{P} = P_1)$ $$\le |P_2(A) - \mu_n(A)| = \mu_n(A) - P_2(A)$$

$$\begin{aligned}\Rightarrow \mathbb{1}_E\, d(P_1, P_2) &\le 2\,\mathbb{1}_E(\mu_n(A) - P_2(A)) \\ &= 2\,\mathbb{1}_E(\mu_n(A) - \bar{P}(A)) \\ &\le 2\,\mathbb{1}_E\, d(\bar{P}, P) + 2|\mu_n(A) - P(A)|\end{aligned}$$

(ii) $\underbrace{\mu_n(A) \geq P_1(A) \geq P_2(A)}$

$$\mathbb{1}_E\, d(P_1, P_2) \leq \mathbb{1}_E \left(\mu_n(A) - P_2(A)\right)$$

$$\leq \mathbb{1}_E\, d(\bar{P}, P) + |\mu_n(A) - P(A)|$$

(iii) $\underbrace{P_1(A) \geq P_2(A) \geq \mu_n(A)}$ Contradicts $\hat{P} = P_1$.

Thus,
$$d(\hat{P}, P) \leq 3\, d(P, \bar{P}) + 2 \max\left\{ |\mu_n(A) - P(A)|, \; |\mu_n(A^c) - P(A^c)| \right\}.$$

$\blacksquare$

Note that $|\mu_n(A) - P(A)|$ can be large for an arbitrary set, but when A has some structure, this deviation can be seen to be small. We will revisit this later, but we first extend the result above to a eneral $M \geq 2$.

We present a selector that uses $O(M^2)$ computations of the form $P_i(A)$.

<u>Scheffé Tournament</u>: Compare each pair (P_i, P_j), $1 \leq i < j \leq M$ and declare the winner as that i which wins the maximum number of matches. Denote the winner by $\hat{P}$.

<u>Theorem</u>
$$d(\hat{P}, P) \leq 9 \min_i d(P_i, P) + 8\Delta$$

where
$$\Delta = \max_{A \in \{A(f_i, f_j) : 1 \leq i < j \leq M\}} |P(A) - \mu_n(A)|.$$

Proof. Let $a = \min_i d(P_i, P)$ and denote

$\mathcal{M}_1 = \{i : a < d(P_i, P) \le 3a + 2\Delta\}$

$\mathcal{M}_2 = \{i : 3a + 2\Delta < d(P_i, P) \le 9a + 8\Delta\}$

$\mathcal{M}_3 = \{i : d(P_i, P) > 9a + 8\Delta\}$.

Then, any i s.t. $d(P_i, P) = a$ must win against those in $\mathcal{M}_2$ and $\mathcal{M}_3$. Next consider $i \in \mathcal{M}_3$ and j s.t. $d(P_j, P) = a$ or $j \in \mathcal{M}_1$. Then,

$$d(\hat{P}_{ij}, P) \le 3\, d(P_j, P) + 2\Delta \le 9a + 8\Delta,$$

which implies $\hat{P}_{ij} = P_j$. Thus, any member of $\mathcal{M}_3$ can win at most $|\mathcal{M}_2| + |\mathcal{M}_3| - 1$ matches while a member of $\mathcal{M}_1$ or the minimizer wins $|\mathcal{M}_2| + |\mathcal{M}_3|$ matches. Thus, a member of $\mathcal{M}_3$ can never win the tournament. ∎