Lecture 7

Review:

* Scheffé Selector

Let $P_1, ..., P_m$ be measures on $\mathcal{X}$. Given samples $X_1, ..., X_n$ from $P$, find i.s.t.

$$d(P_i, P) \leq c \cdot \min_i d(P_i, P).$$

→ Scheffé selector for $M = 2$

Let $A = \{ x : f_1(x) \geq f_2(x) \}$.

$$\hat{P} = \begin{cases} P_i, & \text{if } |P_i(A) - \mu_n(A)| \leq |P_2(A) - \mu_n(A)|. \\ P_2, & \text{o.w.} \end{cases}$$

→ Scheffé Tournament

- Use Scheffé selector to find a winner for each "match" $P_i$ vs $P_j$, $1 \leq i < j \leq M$.
- Choose the winner as the $P_i$ that wins the most matches.

→ Performance of Scheffé tournament

$$d(\hat{P}, P) \leq 9 \min_i d(P_i, P) + 8 \Delta$$

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

Agenda:

Estimating the mean of a Gaussian using Scheffé selector
We modify the problem a little bit (we shall illustrate in HW that this modification is without loss of generality).

\[ e(n, s) = \min_{\hat{\mu}} \max_{\mu} \mathbb{P}_\mu \left( d(\hat{\mu}, \mu) > s \right). \]

**New recipe**

(i) Form a list of guesses \( \mu_1, ..., \mu_m \) using \( X^n \)

(ii) Use a Scheffe selector to find \( \hat{\mu} = \mu_i \).

Recall that

\[ d(\mathbb{P}_\mu, \mathbb{P}_\nu) \leq \frac{1}{2} |\mu - \nu| \]

**Step (i)**

Consider \( n \) independent samples \( X_1, ..., X_n \) from \( N(\mu, \sigma^2) \).

Then,

\[ P \left( |X_i - \mu| > t \right) = 2 \Phi \left( \frac{t}{\sigma} \right) \leq 2 e^{-\frac{t^2}{2\sigma^2}} \]

Thus,

\[ P \left( \min_i |X_i - \mu| > t \right) = P \left( |X_i - \mu| > t \right)^n \]

\[ \leq 2 e^{-\frac{n t^2}{2\sigma^2}} = \frac{\varepsilon}{2} \]

if \( t = \frac{2\sigma}{\sqrt{n} \log \frac{4}{\varepsilon}} \).

i.e., with prob. greater than \( 1 - \varepsilon/2 \), there exists \( i \) s.t.

\[ |X_i - \mu| \leq \frac{2\sigma}{\sqrt{n}} \log \frac{4}{\varepsilon} \]

Thus, a simple guess list is \( \{ X_1, ..., X_n \} \).
Step(ii) We now use Scheffe tournament to find a good
candidate from the guess-list. The selected \( \hat{P} \) satisfies

\[
d(\hat{P}, P) \leq \min_{1 \leq i \leq n} d(P_i, P) + 8\Delta
\leq \frac{9}{\sqrt{n}} \log \frac{4}{\delta} + 8\Delta
\]

with prob. \( \geq 1 - \epsilon/2 \).

Here, in the definition of \( \Delta \), the set

\[
A(\mu_i, \mu_j) = \{ x : |x - \mu_i| \leq |x - \mu_j| \} = \{ x : x \in \frac{\mu_i + \mu_j}{2} \}.
\]

Thus,

\[
\Delta = \max_{i, j} \left| P_\mu \left( X \leq \frac{\mu_i + \mu_j}{2} \right) - P_\mu \left( X \leq \frac{\mu_i + \mu_j}{2} \right) \right|
\leq \max_{a \in \mathbb{R}} \left| P_\mu \left( X \leq a \right) - P_\mu \left( X \leq a \right) \right|
\]

Convergence of Empirical measure to the original one

Let \( \mu_n \) denote the empirical measure generated using \( n \) samples
from \( \mu \) on \( \mathbb{R} \).

Then,

\[
\mu \left( \sup_{a} \left| \mu (X \leq a) - \mu_n (X \leq a) \right| > \frac{1}{\sqrt{n}} + t \right) \leq e^{-nt^2/2}
\]
Using (11),

\[ P_M \left( d(M, P) > \frac{9 \sigma}{\sqrt{n}} \log \frac{4}{\epsilon} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sqrt{2 \log \frac{2}{\epsilon}} \right) \leq \epsilon. \]

Thus, \( n\epsilon(S) \leq O\left( \frac{C^2}{\delta} \log \frac{1}{\epsilon} \right) \)

**Proof of (11) (See lectures 6, 7 of EE2207)**

We use a result of Massart '90:

**Kolmogorov-Smirnov Distance (for \(\mu, \nu\) on \(\mathbb{R}\))**

\[ d_{KS}(\mu, \nu) = \sup_{a \in \mathbb{R}} |\mu(X \leq a) - \nu(X \leq a)|. \]

**Massart '90:** \[ E_{\mu} \left[ d_{KS}(\mu, \mu_n) \right] \leq \frac{1}{\sqrt{n}}. \]

Note that the function \( d_{KS}(\mu, \mu_n) \) satisfies BDD with constants \( (K_n, \ldots, K_n) \). Thus, by Mc Diarmid's inequality

\[ \mu \left( d_{KS}(\mu_n, \mu) > E_{\mu} \left[ d_{KS}(\mu_n, \mu) \right] + t \right) \leq e^{-nt^2} \]

\[ \Rightarrow \mu \left( d_{KS}(\mu_n, \mu) > \frac{1}{\sqrt{n}} + t \right) \leq e^{-nt^2}. \]

The same analysis can be extended to higher dimensions. The queries would be \( X_1, \ldots, X_n \) and then the Scheffe selector can return the best \( i \) to be used as a proxy for \( \mu_i \).
But now we will need $O(d \log d)$ samples and the Scheffé selector will work in $O(d^2 \log^2 d)$ steps. This is "comparable" with the $O(d^3)$ steps required for calculating the empirical mean of $O(d)$ samples.