Lecture 7

Review: * Scheffé Selector

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

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

→ Scheffé selector for $M=2$

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

$$\hat{P} = \begin{cases} P_1, & \text{if } |P_1(A) - \mu_n(A)| \le |P_2(A) - \mu_n(A)|, \\ P_2, & \text{o.w.} \end{cases}$$

→ Scheffé Tournament

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

→ Performance of Scheffé tournament

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

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

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).

$$\epsilon(n,\delta) = \min_{\hat{\mu}} \max_{\mu} P_\mu\left(d\left(P_\mu, P_{\hat{\mu}}\right) > \delta \right).$$

New recipe

(i) Form a list of guesses $\mu_1, \dots, \mu_M$ using X^n

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

Recall that

$$d\left(P_\mu, P_\nu\right) \le \frac{1}{2} |\mu - \nu|.$$

Step (i)

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

Then, $P\left(|X_i - \mu| > t \right) = 2Q\left(\frac{t}{\sigma}\right) \le 2 \cdot e^{-t^2/2\sigma^2}$

Thus,

$$P\left(\min_i |X_i - \mu| > t \right) = P\left(|X_1 - \mu| > t \right)^n$$

$$\le 2 \cdot e^{-nt^2/2\sigma^2} = \frac{\epsilon}{2}$$

if $t = \frac{2\sigma}{\sqrt{n}} \log \frac{4}{\epsilon}$.

i.e., with prob. greater than $1 - \epsilon/2$, there exists i s.t.

$$|X_i - \mu| \le \frac{2\sigma}{\sqrt{n}} \log \frac{4}{\epsilon}.$$

Thus, a simple guess list is $\{P_{X_1}, \dots, P_{X_n}\}$.

Step(ii) We now use Scheffé tournament to find a good candidate from the guess-list. The selected $\hat{P}$ satisfies

$$d(\hat{P}, P) \le 9 \min_{1 \le i \le n} d(P_i, P) + 8\Delta$$

$$\le \frac{9\sigma}{\sqrt{n}} \log \frac{4}{\epsilon} + 8\Delta$$

with prob. $\ge 1 - \epsilon/2$.

Here, in the definition of Δ, the set

$$A(\mu_i, \mu_j) = \{x : |x - \mu_i| \le |x - \mu_j|\} = \{x : x \le \frac{\mu_i + \mu_j}{2}\}.$$

Thus,

$$\Delta = \max_{i,j} \left| P_\mu\left(X \le \frac{\mu_i + \mu_j}{2}\right) - \mu_n\left(X \le \frac{\mu_i + \mu_j}{2}\right) \right|$$

↳ excuse the bad notation!!

$$\le \max_{a \in \mathbb{R}} \left| P_\mu(X \le a) - \mu_n(X \le a) \right|$$

Convergence of Empirical measure to the original one

Let μ_n denote the empirical measure generated using n samples from μ on $\mathbb{R}$.

Then,

$$(1) \quad \mu\left(\sup_a |\mu(X \le a) - \mu_n(X \le a)| > \frac{1}{\sqrt{n}} + t \right) \le e^{-nt^2/2}$$

Using (1),

$$P_\mu\left(d(\hat{P},P) > \frac{9\sigma}{\sqrt{n}} \log\frac{4}{\epsilon} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sqrt{2\log\frac{2}{\epsilon}} \right) \le \epsilon .$$

Thus, $n_\epsilon(\delta) \le O\left(\frac{\sigma^2}{\delta} \log\frac{1}{\epsilon}\right)$

Proof of (1) (See lectures 6, 7 of E2207)

We use a result of Massart '90:

Kolmogorov-Smirnov Distance (for μ, ν on $\mathbb{R}$)

$$d_{KS}(\mu,\nu) = \sup_{a\in\mathbb{R}} \left|\mu(X\le a) - \nu(X\le a)\right| .$$

Massart '90: $\mathbb{E}_\mu\left[d_{KS}(\mu,\mu_n)\right] \le \frac{1}{\sqrt{n}}$.

Note that the function $d_{KS}(\mu,\mu_n)$ satisfies BDD with constants $(1/n, \dots, 1/n)$. Thus, by McDiarmid's inequality

$$\mu\left(d_{KS}(\mu_n,\mu) > \mathbb{E}_\mu[d_{KS}(\mu_n,\mu)] + t\right) \le e^{-nt^2/2}$$

$$\Rightarrow \mu\left(d_{KS}(\mu_n,\mu) > \frac{1}{\sqrt{n}} + t\right) \le e^{-nt^2/2} . \quad \blacksquare$$

The same analysis can be extended to higher dimensions. The guesses would be $\{X_1, \dots, X_n\}$ and then the Scheffe selector can return the best i to be used as a proxy for μ_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 $O(d^2)$ steps required for calculating the empirical mean of $O(d)$ samples.