Lecture 3

Review: – Total variation distance

$$* \quad d(P,Q) \triangleq \sup_A P(A) - Q(A) = \frac{1}{2}\sum_x |P(x)-Q(x)|$$

$$* \quad d(P^n, Q^n) \le n\, d(P,Q)$$

– KL divergence

$$* \quad D(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

$*$ Pinsker's inequality: $d^2(P,Q) \le \frac{1}{2\ln 2} D(P\|Q)$

$$* \quad d(P^n, Q^n) \le \sqrt{\frac{n D(P\|Q)}{2\ln 2}}$$

– Fano's inequality.

$*$ Consider an M-ary hypothesis testing problem.

$$P_e \geq 1 - \frac{\frac{1}{M}\sum_{m=1}^{M} D\left(P_m \| \frac{1}{M}\sum_{m'} P_{m'}\right) + h(P_e)}{\log M}$$

$$\Rightarrow P_e \geq 1 - \frac{\max_{m,m'} D(P_m\|P_{m'}) + h(P_e)}{\log M}$$

For binary HT: $P_e \geq (1-h(P_e)) - D(P^n\|Q^n)$

$$\Rightarrow nD(P\|Q) \geq 1 - P_e - h(P_e) > 1/3 \text{ for } p \text{ small}$$

$\boxed{A}$ Estimating the bias of a coin

The minmax framework

$$R_n = \min_{\hat{p}} \max_{p\in[0,1]} \mathbb{E}_p\left[(p-\hat{p}(X^n))^2\right]$$

The "probably approximately correct" (PAC) framework

$$\epsilon(\delta, n) = \min_{\hat{p}} \max_{p \in [0,1]} \mathbb{P}_p\left(|p - \hat{p}(x^n)| > \delta \right)$$

Sample complexity: $n_\epsilon(\delta) = \min\{n : \epsilon(n, \delta) \leq \epsilon\}$

(1) The estimator: $\hat{p}(x^n) = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$* \ \hat{p}(\underline{x}) = \arg\max_{p} \ p^{k(\underline{x})}(1-p)^{n-k(\underline{x})}$$

$$= \arg\max_{p} \ 2^{n\left(\frac{k(\underline{x})}{n}\log p + \frac{n-k(\underline{x})}{n}\log\frac{1}{1-p}\right)}$$

$$= \arg\max_{p} \ 2^{-n\left(\hat{p}\log\frac{\hat{p}}{p} + (1-\hat{p})\log\frac{1-\hat{p}}{1-p}\right)}$$

$$= \arg\max_{p} \ 2^{-n D(\hat{p}\|p)} = \hat{p}$$

$$\Rightarrow \boxed{\hat{p} \text{ is the ML estimate}}$$

Note $\mathbb{E}_p\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right)^2\right]$

$$= \frac{p(1-p)}{n} \leq \frac{1}{4n}$$

Thus, $R_n \leq \frac{1}{4n}$.

Using Markov's inequality

$$\mathbb{P}_p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \delta \right) \leq \frac{1}{\delta\sqrt{4n}}$$

$$\Rightarrow n_\epsilon(\delta) \leq \frac{4}{\delta^2}\cdot\frac{1}{\epsilon^2}.$$

(2) The lower bound (Fano's inequality)

Consider $p, q \in [0,1]$. Then, for $P_e = \frac{1}{2} P_{e,p} + \frac{1}{2} P_{e,q}$,

$$n\, D(p \| q) \geq 1 - P_e - h(P_e).$$

But how do we use it to get a lower bound for $n_\epsilon(\delta)$?

Reduction of estimation to testing

(i) Minmax cost. For any estimator $\hat{p}$,

$$\max_p \mathbb{E}_p\left[(p - \hat{p}(X))^2\right]$$

$$\geq \max\left\{ \mathbb{E}_p\left[(p - \hat{p}(X))^2\right], \mathbb{E}_q\left[(q - \hat{p}(X))^2\right]\right\}$$

Let
$$e(\hat{p}) = \begin{cases} p, & \text{if } |p - \hat{p}| \leq |p-q|/2, \\ q, & \text{if } |q - \hat{p}| < |p-q|/2. \end{cases}$$

Then,
$$\mathbb{P}_p(e(\hat{p}) = q) \leq \mathbb{P}_p\left(|p - \hat{p}| > |p-q|/2\right)$$
$$\leq \frac{4\,\mathbb{E}_p\left[|p-\hat{p}|^2\right]}{(p-q)^2}$$

$$\Rightarrow R_n \geq \frac{(p-q)^2}{4} \max\{P_{e,p}; P_{e,q}\}$$
$$\geq \frac{(p-q)^2}{4} \frac{1}{2}\left(P_{e,p} + P_{e,q}\right)$$
$$\geq \frac{(p-q)^2}{4} P_e$$

Similarly, for any $p_1, \dots, p_m$ s.t. $|p_i - p_j| \geq \delta$,
$$R_n \geq \frac{\delta^2}{4} \cdot P_e^*(\text{unif})$$

For $M > 2$, we can simply use

$$P_e^* \geq 1 - \frac{\max\limits_{i,j} D(P_i \| P_j) + 1}{\log M}$$

But for $M = 2$, we can use

$$P_e^* = \frac{1}{2}\left(1 - d(P^n, Q^n)\right)$$

$$\geq \frac{1}{2}\left(1 - \sqrt{n D(p\|q)}\right)$$

Thus, $R_n \geq \frac{(p-q)^2}{8}\left(1 - \sqrt{n D(p\|q)}\right)$

Choose $p = \frac{1}{2}$, $q = \frac{1}{2} + \delta$

$$R_n \geq \frac{\delta^2}{8}\left(1 - \sqrt{n D\left(\tfrac{1}{2} \| \tfrac{1}{2} + \delta\right)}\right)$$

$$D\left(\tfrac{1}{2} \| \tfrac{1}{2} + \delta\right) = 1 - h\left(\tfrac{1}{2} + \delta\right)$$

$$h'(x) = \ln\left(\frac{1-x}{x}\right), \quad h''(x) = \frac{-1}{x(1-x)}$$

$$\Rightarrow h'\left(\tfrac{1}{2}\right) = 0$$

$$\Rightarrow 1 - h\left(\frac{1}{2} + \delta\right) \leq \max_{\frac{1}{2} \leq x \leq \frac{1}{2} + \delta} \frac{1}{x(1-x)} \delta^2$$

$$\leq \frac{16}{3}\delta^2 \qquad \left(\text{assuming } \delta \leq \tfrac{1}{4}\right)$$

Therefore, $R_n \geq \frac{\delta^2}{8}\left(1 - \delta\sqrt{\frac{16n}{3}}\right)$

Choose $\delta = \frac{1}{2}\sqrt{\frac{3}{16n}} \Rightarrow \boxed{R_n \geq \frac{3}{1024} \cdot \left(\frac{1}{n}\right)} \Rightarrow R_n = \Theta\left(\frac{1}{n}\right)$

(ii) PAC framework

A similar method can be used:

$$\frac{1}{2}\left(\mathbb{P}_p\left(|p-\hat{p}(X^n)|>\delta\right)+\mathbb{P}_q\left(|q-\hat{p}(X^n)|>\delta\right)\right)$$

$$\geq P_e^*\left(\frac{1}{2},\frac{1}{2}\right)$$

$$= \frac{1}{2}\left(1-\sqrt{n\,D(p\|q)}\right)$$

$$\geq \frac{1}{2}\left(1-\delta\sqrt{\frac{16n}{3}}\right)$$

$$\Rightarrow n_\epsilon(\delta) \geq \frac{3.(1-2\epsilon)}{16.\,\delta^2} = \Omega\left(\frac{1}{\delta^2}\right).$$

Thus, $n_\epsilon(\delta) = \Theta_\epsilon\left(\frac{1}{\delta^2}\right)$ [We will see $n_\epsilon(\delta) = O\left(\frac{1}{\delta^2}\log\frac{1}{\epsilon}\right)$]