Lecture 4

Review

* $R_n = \min_{\hat\theta} \max_{\theta} \mathbb{E}_\theta [r(\theta, \hat\theta(X^n))]$ minmax risk

$\epsilon(n,\delta) = \min_{\hat\theta} \max_{\theta} \mathbb{P}_\theta \left(r(\theta, \hat\theta(X^n)) > \delta \right)$ PAC cost

$n_\epsilon(\delta) = \min \{ n \mid \epsilon(n,\delta) \le \epsilon \}$ sample complexity

* Lower bounds (Θ, d) be a metric space.

Consider $\theta_1, \dots, \theta_M$ s.t. $\min_{i,j} d(\theta_i, \theta_j) \ge 2\delta$

$$R_n \ge \delta^2 \underbrace{P_e^*(\text{unif} \mid \theta_1, \dots, \theta_M)}_{\text{min } P_e \text{ for } M\text{-ary HT with uniform prior}}$$

→ Le Cam's two point method $\boxed{M=2}$

$$P_e^*(\text{unif}) \ge \frac{1}{2}\left(1 - \sqrt{n D(P_{\theta_1} \| P_{\theta_2})}\right)$$

→ Fano's inequality

$$P_e^*(\text{unif}) \ge \left(1 - \frac{\max_{i,j} D(P_{\theta_i} \| P_{\theta_j}) + 1}{\log M}\right)$$

* For the coin toss example,

$$n_\epsilon(\delta) = \Theta_\epsilon\left(\frac{1}{\delta^2}\right); \quad R_n = \Theta\left(\frac{1}{n}\right)$$

Agenda

* Concentration inequalities
* Median trick
* Estimating Gaussian mean

$\boxed{A}$ Concentration inequalities

We now present bounds that will enable us to show

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

Markov inequality $X \geq 0$

$$P(X > t) \leq \frac{t}{\mathbb{E}[X]}$$

Chebyshev inequality

$$P(|X - \mathbb{E}X| > t) \leq \frac{t^2}{\mathrm{Var}(X)}$$

Chernoff bound

$$P(X > \mathbb{E}X + t) = P\left(e^{\lambda(X - \mathbb{E}X)} > e^{\lambda t}\right)$$

$$\leq e^{-\lambda t} \cdot \mathbb{E}\left[e^{\lambda(X - \mathbb{E}X)}\right]$$

$$= e^{\psi_{X-\mathbb{E}X}(\lambda) - \lambda t}$$

$$\psi_{X-\mathbb{E}X}(\lambda) = \log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}X)}\right]$$: log-moment generating function

Theorem (Hoeffding's inequality)

Let $X_1, \dots, X_n$ be indep. and such that $X_i \in [a_i, b_i]$.

$$P\left(\sum_{i=1}^n X_i > t + \sum_{i=1}^n \mathbb{E}[X_i]\right) \leq e^{-2t^2 / \sum_{i=1}^n (b_i - a_i)^2}$$

We will prove a much stronger bound.

Let $\Delta_1, \dots, \Delta_n$ be zero mean, uncorrelated, i.e.,

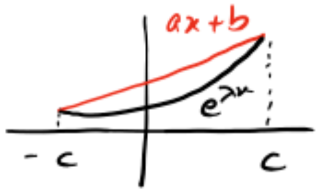
$$\mathbb{E}[\Delta_i] = 0, \quad \mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\Delta_i]\,\mathbb{E}[\Delta_j], \quad i \neq j.$$

Further, let Δ_i be s.t. $|\Delta_i| \leq c_i$, $1 \leq i \leq n$.

Then,
$$\boxed{\mathbb{P}\left(\sum_{i=1}^n \Delta_i^2 > t\right) \leq \exp\left(-t^2 / 2\sum_{i=1}^n c_i^2\right)}$$

Proof:

For $\Delta_1, \dots, \Delta_n$ as above:



$$\mathbb{E}\left[e^{\lambda \sum_i \Delta_i}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{\lambda \Delta_i}\right]$$

$$e^{\lambda x} \le \frac{e^{\lambda c} - e^{-\lambda c}}{2} \cdot x + \frac{e^{\lambda c} + e^{-\lambda c}}{2}, \quad -c \le x \le c$$

Thus, $\mathbb{E}\left[\prod_{i=1}^n e^{\lambda \Delta_i}\right] \le \mathbb{E}\left[\prod_{i=1}^n (a_i X_i + b_i)\right]$

$$\le \prod_{i=1}^n b_i$$

$$= \prod_{i=1}^n \frac{e^{-\lambda c_i} + e^{\lambda c_i}}{2}$$

$$= \prod_{i=1}^n \left(1 + \frac{\lambda^2 c_i^2}{2!} + \frac{\lambda^4 c_i^4}{4!} + \cdots\right)$$

$$\le \prod_{i=1}^n e^{\lambda^2 c_i^2/2} = e^{\lambda^2 \sum_{i=1}^n c_i^2/2}$$

Thus, by Chernoff bound,

$$P\left(\sum_{i=1}^n \Delta_i > t\right) \le \exp\Big(\underbrace{\frac{\lambda^2}{2}\sum_{i=1}^n c_i^2 - \lambda t}_{g(\lambda)}\Big)$$

$g(\lambda) = \frac{\lambda^2}{2} c - \lambda t$

$\max_{\lambda > 0} = -t^2/2$ if $t > 0$

Thus, $P\left(\sum_{i=1}^n \Delta_i > t\right) \le \exp\left(-t^2/2 \sum_{i=1}^n c_i^2\right)$ ∎

We would like to use this concentration bound to get concentration for a function $f(X^n)$

Let $\Delta_i = \mathbb{E}[f(X^n)|X^i] - \mathbb{E}[f(X^n)|X^{i-1}]$.

Then, (a) $\sum_{i=1}^{n} \Delta_i = f(X^n) - \mathbb{E}[f(X^n)]$

(b) $\mathbb{E}[\Delta_i] = 0$

(c) For $i > j$, $\mathbb{E}[\Delta_i \Delta_j] = \mathbb{E}[\mathbb{E}[\Delta_i \Delta_j | X^j]]$

$$= \mathbb{E}[\Delta_j \, \mathbb{E}[\Delta_i | X^j]]$$

$$\mathbb{E}[\Delta | X^j] = \mathbb{E}[\mathbb{E}[f(X^n)|X^i]|X^j] - \mathbb{E}[\mathbb{E}[f(X^n)|X^{i-1}]|X^j] = 0.$$

<u>Definition</u> $f: \mathcal{X}^n \to \mathbb{R}$ satisfies the <u>bounded difference property</u> (BDP) with constants $c = (c_1, \dots, c_n)$ if

$$f(x) - f(y) \le c_i \quad \forall\, x, y \text{ s.t. } x_j = y_j \text{ for } j \ne i.$$

For a function f satisfying BDP with c, $|\Delta_i| \le c_i$.

<u>Theorem</u> (McDiarmid's inequality)

$X_1, \dots, X_n$ be indep.

f satisfies BDP with c

$$P(f(X) > \mathbb{E}f(X) + t) \le \exp\left(- t^2 / 2 \sum_{i=1}^{n} c_i^2\right).$$

<u>Remarks</u>: (a) <u>Chebyshev's ineq.</u> $\quad v = \mathrm{Var}(f(X))$

$$P\left(f(X) > \mathbb{E}f(X) + \sqrt{\tfrac{v}{\epsilon}}\right) \le \epsilon$$

(b) McDiarmid's ineq. Let f satisfy BDP with c

$$P\Big(f(X) \geq \mathbb{E} f(X) + \sqrt{\underbrace{2\sum_{i=1}^{n} c_i^2}_{\text{variance parameter}} \log 1/\epsilon} \Big) \leq \epsilon.$$

Returning to the coin-toss example,

$$\mathbb{P}_p\Big(\Big|\frac{1}{n}\sum_i X_i - p\Big| > \delta \Big) \leq 2 \exp\Big(-\frac{2n^2\delta^2}{2n}\Big)$$
$$= 2e^{-n\delta^2}$$

$$\Rightarrow n_\epsilon(\delta) \leq \frac{1}{\delta^2} \log \frac{2}{\epsilon}$$

$\boxed{B}$ Median Trick

We now present a neat trick that will allow us to restrict to $\epsilon = 1/3$ for problems with 1-dimensional parameter space.

Lemma (Median trick) Let $\Theta \subseteq \mathbb{R}$.

Suppose that we have an estimator $\hat{\theta}$ s.t.

$$P_\theta\big(|\theta - \hat{\theta}(X)| > \delta \big) \leq \frac{1}{3}.$$

Then, we can find another estimator $\hat{\theta}$ s.t.

$$P_\theta\big(|\theta - \hat{\theta}(X^n)| > \delta \big) \leq \epsilon,$$

where $X_1, \ldots, X_n$ are iid with $n = O\big(\log \frac{1}{\epsilon}\big)$.

Proof. Let $\hat{\theta}(X^n) = \text{med}\{\hat{\theta}(X_1), \ldots, \hat{\theta}(X_n)\}$.

Then, $|\hat{\theta}(X^n) - \theta| > \delta \Rightarrow |\{i : |\hat{\theta}(X_i) - \theta| > \delta\}| \geq \frac{n}{2}$.

Thus, $P_\theta\left(|\hat{\theta}(X^n)-\theta| > \delta\right) \le P\left(\sum_{i=1}^{n} B_i \ge \frac{n}{2}\right)$

where $B_1, \dots, B_n$ are iid $\mathrm{Ber}(p)$, with $p \le \frac{1}{3}$.

Using Hoeffding's inequality,

the right-side $\le e^{-2\left(\frac{n}{2}-\frac{n}{3}\right)^2 \frac{1}{n}} = e^{-\frac{n}{18}} \le \epsilon$

if $n \ge 18 \ln \frac{1}{\epsilon}$. ∎

$\boxed{C}$ Estimating the Gaussian mean

$X_1, \dots, X_n \sim N(\mu, \sigma^2 I_{d\times d})$ are values in $\mathbb{R}^d$

$$R_n = \inf_{\hat{\mu}} \sup_{\mu \in \mathbb{R}^d} \mathbb{E}_\mu\left[\|\hat{\mu}(X^n) - \mu\|_2\right]$$

The Estimator

$$\hat{\mu}(X^n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, $\mathbb{E}_\mu\left[\left(\hat{\mu}_i(X^n) - \mu_i\right)^2\right] = \frac{1}{n^2} \cdot \mathbb{E}_{\mu_i}\left[\left(\sum_{t=1}^{n} X_{t,i} - n\mu_i\right)^2\right]$

$= \frac{\sigma^2}{n} \Rightarrow R_n \le \frac{d\sigma^2}{n}$

The Lower Bound

Consider $\mu_1, \dots, \mu_M$ s.t. $\|\mu_i\|_2 \le 1$ and s.t.

$$\min_{i \ne j} \|\mu_i - \mu_j\|_2 \ge 2\delta.$$

Then, $R_n \ge \delta^2 P_e^*(\mathrm{unif}\,|\,\mu_1, \dots, \mu_M)$

Le Cam's Method $\quad \|\mu - \nu\|_2 = \delta$

Then,
$$R_n \geq \frac{\delta^2}{2}\left(1 - \sqrt{n\, D(P_\mu \| P_\nu)}\right)$$

Note that $D(P_\mu \| P_\nu)$
$$= \frac{1}{2\sigma^2} \mathbb{E}_\mu\left[\sum_{i=1}^{d} (X_i - \nu_i)^2 - (X_i - \mu_i)^2\right]$$
$$= \frac{1}{2\sigma^2} \sum_{i=1}^{d} \mathbb{E}_\mu\left[\nu_i^2 - \mu_i^2 - 2X_i(\nu_i - \mu_i)\right]$$
$$= \frac{1}{2\sigma^2} \sum_{i=1}^{d} \nu_i^2 + \mu_i^2 - 2\mu_i\nu_i = \frac{\|\mu - \nu\|_2^2}{2\sigma^2} = \frac{\delta^2}{2\sigma^2}$$

Thus, we obtain $R_n \geq \frac{\delta^2}{2}\left(1 - \sqrt{\frac{n\delta^2}{2\sigma^2}}\right)$

Choose $\frac{n\delta^2}{2\sigma^2} = \frac{1}{4} \Rightarrow \delta^2 = \frac{\sigma^2}{2n}$

$\Rightarrow R_n \geq \frac{\sigma^2}{8n}$!! Doesn't grow with d !!