Lecture 10

Last time we were trying prove the following theorem:

Theorem: Let $0<\alpha<1/4$ and $x(\eta)=\int \eta^2\,d\mu$

(i) There exists an estimator $\hat{x}_n$ such that

$$\sup_{\eta\in\mathcal{P}(\alpha,M,B)} \mathbb{E}_\eta(\hat{x}_n - x(\eta))^2 \leq c(M,B)\left(\frac{n^2}{\log n}\right)^{-\frac{4\alpha}{4\alpha+1}}$$

(ii) If for some estimator T_n, one has for some $\alpha<1/4$

$$\sup_{\eta\in\mathcal{P}(\alpha,M,B)} \mathbb{E}_\eta(T_n - x(\eta))^2 \leq c\left(\frac{n^2}{\log n}\right)^{-\frac{4\alpha}{4\alpha+1}}$$

Then $\exists\, \alpha'<\alpha$ s.t.

$$\sup_{\eta\in\mathcal{P}(\alpha',M,B)} \mathbb{E}_\eta\left(T_n - x(\eta)\right)^2 \geq c(M,B)\left(\frac{n^2}{\log n}\right)^{-\frac{4\alpha'}{4\alpha'+1}}$$

Proof of (i):

We have laid down the general idea to prove (i). Then we showed how to obtain a two point adaptation over any $0<\alpha_1<\alpha_0<1/4$. We now use that idea to solve (i).

Fix a $c>1$ and let N be the largest integer s.t. $c^{N-1}\leq n^{1-2/\log\log n}$ i.e. $(N-1)\leq \frac{\log n}{\log c}\left(1-\frac{2}{\log\log n}\right)$

For $\ell=1,\dots,N$, let $2^{j_\ell}\sim c^{\ell-1}n$ and let α_ℓ be the solution to $2^{j_\ell}=n^{\frac{2}{1+4\alpha_\ell}}$ $\Rightarrow k_0<k_1<\dots k_{N-1}$ and $\frac{1}{4}>\alpha_0>\alpha_1>\dots\alpha_{N-1}\geq\frac{1}{4\log\log n-1}$

Also, $\exists\, c_1$ and c_2 s.t. $\alpha_{l_1}-\alpha_{l_2} \in \left[c_1 \frac{l_2-l_1}{\log n}, c_2 \frac{l_2-l_1}{\log n}\right]$ for $l_2>l_1$.

Let j^*_l be s.t. $2^{j^*_l} \sim 2^{j_l} \times (\log n)^{-\frac{1}{1+4\alpha_l}}$

$$\sim \left(\frac{n^2}{\log n}\right)^{1/1+4\alpha_l} \qquad l=0,1,\dots,N-1$$

Indeed $2^{j^*_l}$ increases with l (check this). Finally let $R(j^*_l) = \left(2^{j^*_l}/n^2\right) =$

Let,

$$\hat{l} = \min\left\{ l : \left(\hat{x}_n(j^*_l) - \hat{x}_n(j^*_{l'})\right)^2 \leq C^* \log n \, \frac{2^{j^*_{l'}}}{n^2} \quad \forall\, l' > l \right\}$$

i.e. $\hat{l}$ is the surrogate for first time the $(\text{bias})^2$ goes below the variance.

Finally use the estimator $\hat{x}_n(j^*_{\hat{l}})$. We now analyze the MSE of this estimator & show control at desired rate.

Fix $\alpha \in (0, 1/4)$. Find $l=0,\dots,N-2$ s.t. $\alpha \in (\alpha_{l+1}, \alpha_l]$. Indeed, this l depends on both α and n, and therefore, to be precise we should write $l(\alpha, n)$. However for the sake of brevity we omit such notation.

Now fix $\eta \in \mathcal{P}(\alpha, M, B)$. We need to analyze

$$\mathbb{E}_\eta\left(\hat{x}_n(j^*_{\hat{l}}) - x(\eta)\right)^2 = \mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\hat{l}}) - x(\eta)\right)^2 I(\hat{l} \leq l+1)\right)$$
$$+ \mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\hat{l}}) - x(\eta)\right)^2 I(\hat{l} > l+1)\right)$$
$$= T_1 + T_2$$

Control of T_1: This term corresponds to choosing a smoothness that is not lower than the actual smoothness.

$$\mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\hat{\ell}}) - x(\eta)\right)^2 I\left(\hat{\ell} \le \ell+1\right)\right)$$

$$\le 2\,\mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\hat{\ell}}) - \hat{x}_n(j^*_{\ell+1})\right)^2 I(\hat{\ell} \le \ell+1)\right) + 2\,\mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\ell+1}) - x(\eta)\right)^2\right)$$

$$\le \underbrace{2C^* \log n \, \frac{2^{j^*_{\ell+1}}}{n^2}}_{\text{(by definition of } \hat{j}\text{)}} + \underbrace{2\,c(M,B)\left(\left(2^{-j^*_{\ell+1}\alpha}\right)^4 + \frac{2^{j^*_{\ell+1}}}{n^2}\right)}_{\text{(property of } \hat{x}_n(j^*_{\ell+1})\text{)}}$$

$$\le \left\{2C^* + 2C(M,B)\right\}\left(\frac{n^2}{\log n}\right)^{-\frac{4\alpha_{\ell+1}}{4\alpha_{\ell+1}+1}}$$

$$\le \left\{2C^* + 2c(M,B)\right\} c(c_2) \left(\frac{n^2}{\log n}\right)^{-\frac{4\alpha}{4\alpha+1}}$$

as $\alpha \in (\alpha_{\ell+1}, \alpha_\ell]$ and $\alpha_\ell - \alpha_{\ell+1} \le c_2/\log n$

Control of T_2:

$$\mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\hat{\ell}}) - x(\eta)\right)^2 I\left(\hat{\ell} > \ell+1\right)\right)$$

$$= \sum_{\ell'=\ell+2}^{N-1} \mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\ell'}) - x(\eta)\right)^2 I(\hat{\ell} = \ell')\right)$$

$$\le \sum_{\ell'=\ell+2}^{N-1}\left\{\mathbb{E}_\eta\left(\left(\hat{x}_n(j^*_{\ell'}) - x(\eta)\right)^{2q}\right)\right\}^{\frac{1}{q}}\left\{\mathbb{P}(\hat{\ell} = \ell')\right\}^{\frac{1}{p}}$$

$p, q \ge 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$

Now we use the following lemma to complete the proof of the theorem.

Lemma: If $l' \geq l+2$,

$$\mathbb{P}_\eta(\hat{l} = l') \lesssim C(M,B)/n$$

and $$\mathbb{E}_\eta\left((\hat{x}_n(j^*_{l'}) - x(\eta))^{2q}\right) \lesssim \{C(M,B,q)\}^q \left(\frac{2^{j^*_{l'}}}{n^2}\right)^q$$

if C^* is chosen large enough based on B.

Proof: The proof of this lemma is exactly along the same lines as that of controlling $\mathbb{P}_\eta(\hat{j} = j_1)$ for $\alpha = \alpha_0$ in the last lecture (Lecture 9). Try it by yourself by using the U-statistic deviation inequality and Bernstein's Inequality. You will also need to use Rosenthal's Inequality. For the sake of reference, let me recall these results.

Hoeffding's Inequality Let $X_1, \ldots, X_n$ be independent ~~mean 0~~ random variables with $a_i \leq X_i \leq b_i$ for all $1 \leq i \leq n$. Then, $$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right| > t\right) \leq 3\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Rosenthal's Inequality: Let $X_1, \ldots, X_n$ be independent r.v.'s s.t $E(X_i) = 0$ for all $1 \leq i \leq n$. Let $q \geq 2$ be such that $\mathbb{E}(|X_i|^q) < \infty$ $\forall i \in [1, n]$. Then $\exists$ ~~universal~~ $C(q) > 0$ s.t
$$\mathbb{E}\left(\left|\sum_{i=1}^n X_i\right|^q\right) \leq C(q)\left\{\sum_{i=1}^n \mathbb{E}|X_i|^q + \left(\sum_{i=1}^n \mathbb{E}X_i^2\right)^{q/2}\right\}.$$ (Actually $C(q) \leq \left(\frac{c\,q}{\log(q)}\right)^q$ where c is a universal constant)

Ref: V. V. Petrov (1995): Limit Theorems in Probability Theory.

Proof of (ii): The proof the lower bound i.e. the fact that one necessarily pays a logarithmic price below 1/4 smoothness relies on the following version of constrained risk inequality due to Cai and Low (2011)

Constrained Risk Inequality

The constrained risk inequality is a refined restatement of the minimax lower bound idea that we explored before. In a nutshell, it still says that if your functional takes certain values in ~~part~~ different parts of the distribution space such that you cannot efficiently distinguish/list between these spaces using your data, then you can say something about error of estimating your functional.

More concretely suppose $Z \sim P_\theta$, $\theta \in \Theta$. Let $\hat{Q} = \hat{Q}(Z)$ be an estimator of $Q(\theta) \in \mathbb{R}$ based on Z, having bias $B(\theta) := E_\theta(\hat{Q}(Z)) - Q(\theta)$. Now suppose Θ_0 and Θ_1 be a disjoint partition of Θ with priors π_0 and π_1 supported on them.

Let $\mu_i = \int Q(\theta)\, d\pi_i(\theta)$

$$\sigma_i^2 = \int (Q(\theta) - \mu_i)^2 \, d\pi_i(\theta) \qquad i = 0, 1$$

Let γ_i be the marginal ~~distribution~~ density Z under θ. w.r.t some common dominating measure

Let the χ^2-divergence between γ_0 and γ_1 be

$$\chi(\gamma_0, \gamma_1) = \left\{ \mathbb{E}_{\gamma_0} \left(\frac{\gamma_1}{\gamma_0}(Z) - 1 \right)^2 \right\}^{\frac{1}{2}}$$

Then we have the following result.

Lemma (Cai & Low (2011)) If $\int \mathbb{E}_\theta \left(\hat{Q}(Z) - Q(\theta)\right)^2 d\pi_0(\theta) \leq \varepsilon^2$, then

$$\left| \int B(\theta)\, d\pi_1(\theta) - \int B(\theta)\, d\pi_0(\theta) \right| \geq |\mu_1 - \mu_0| - (\varepsilon + \sigma_0)\chi$$

Remark: since maximum risk is always as large as average risk, this will yield a ~~board~~ lower bound on the minimax risk over Θ_1.

Application to prove (ii):

Let $H : [0,1] \to \mathbb{R}$ be a C^∞ function supported on $[0, \frac{1}{2}]$ such that $\int H(x)dx = 0$ and $\int H^2(x)\,dx = 1$, and for $k \in \mathbb{N}$ (to be decided later) take the translates of the interval $\frac{1}{k}[0, \frac{1}{2}]$ that are disjoint and contained in $[0,1]$. Let $x_1, \ldots, x_k$ be the bottom left corners of these cube and $\lambda = (\lambda_1, \ldots, \lambda_k) \in \{-1,+1\}^k$. Let

$$\eta_{\lambda,\alpha'}(x) = 1 + c_0 \left(\frac{1}{k}\right)^{\alpha'} \sum_{j=1}^{k} \lambda_j\, H((x - x_k)k) \qquad \lambda = (\lambda_1 \ldots \lambda_k) \in \{-1,+1\}^k$$

$$\alpha < \frac{1}{4}$$

As we have seen before, a properly chosen $H_{,\lambda}^{c_0}$ guarantees $\eta_{\lambda,\alpha'} \in \mathcal{P}(\alpha', M, B)$. Let

$$\Theta_0^{(n)} = \{P^n : P \equiv U[0,1]\}$$

$$\Theta_1^{(n)} = \{P^n : P = \eta_{\lambda,\alpha'},\ \lambda \in \{-1,+1\}^k\}$$

Let $\Theta^{(n)} = \Theta_0^{(n)} \cup \Theta_1^{(n)}$. $\phi : \Theta^{(n)} \to \mathbb{R}$, $\phi(P^n) = \int \left(\frac{dP}{d\lambda}\right)^2 d\lambda$

Now, $\phi \equiv 1$ on $\Theta_0^{(n)}$. $\Rightarrow$ For any prior π_0 on $\Theta_0^{(n)}$,

$\mu_0 = \int \phi(P^n)\, d\pi_0(P^n) = 1$ and $\sigma_0 = 0$.

Also, $\phi \equiv C(H, G_0)\left(\frac{1}{K}\right)^{\alpha'} + 1$ on $\Theta_1^{(n)}$ for any prior π_1

$\Rightarrow \mu_1 = C(H, G_0)\left(\frac{1}{K}\right)^{\alpha'} + 1$, $\sigma_1 = 0$.

Now, $\Theta_0^{(n)} \subseteq \{P^n : P \in \mathcal{P}(\alpha, M, B)\}$ $\alpha > \alpha'$.

$\Rightarrow$ For any estimator $\hat{\chi}_n$, the π_0 average bias over $\Theta_0^{(n)}$ is bounded by $\varepsilon = C\left(\frac{n}{\sqrt{\log n}}\right)^{-\frac{4\alpha}{4\alpha+1}}$. Therefore by the lemma,

the π_1-average bias of $\hat{\chi}_n$ over $\Theta_1^{(n)}$ must be ~~be~~ at least as large as

$$|\mu_1 - \mu_0| - \left(C\left(\frac{n}{\sqrt{\log n}}\right)^{-\frac{4\alpha}{4\alpha+1}} + 0\right)\chi\left(\nu_0^{(n)}, \nu_1^{(n)}\right)$$

where $\nu_0^{(n)}$ and $\nu_1^{(n)}$ are marginal densities of $(X_1, \ldots, X_n)$ under π_0 and π_1

$$= C(H, G_0)\left(\frac{1}{K}\right)^{\alpha'} - C\left(\frac{n}{\sqrt{\log n}}\right)^{-\frac{4\alpha}{4\alpha+1}} \chi\left(\nu_0^{(n)}, \nu_1^{(n)}\right)$$

Now $\chi^2\left(\nu_0^{(n)}, \nu_1^{(n)}\right)$

$$= \mathbb{E}_{\nu_0^{(n)}}\left(\frac{\nu_1^{(n)}}{\nu_0^{(n)}} - 1\right)^2$$

$$= \mathbb{E}_{P_0^n}\left(\frac{1}{2^k}\sum_{\lambda} \frac{dP^n_{\eta_{\lambda,\alpha'}}}{dP_0^n} - 1\right)^2$$

$P_0^n \leftrightarrow$ product of $U[0,1]$

$P^n_{\eta_{\lambda,\alpha'}} \leftrightarrow$ product of $P_{\eta_{\lambda,\alpha'}}$

This we have learned to control in Lecture 6.

Choosing $k = \left(\frac{n}{\sqrt{c^* \log n}}\right)^{\frac{2}{2\alpha'+1}}$, we get

$\chi^2\left(\nu_0(n), \nu_1(n)\right) \lesssim n^{c'c^*} - 1$ for some constant $c' > 0$.

Now this means maximum bias ~~risk~~ over $\Theta_1^{(n)}$ is at least

$$c(H, c_0)\left(\frac{n}{\sqrt{\log n}}\right)^{-\frac{4\alpha'}{4\alpha'+1}} - c\left(\frac{n}{\sqrt{\log n}}\right)^{-\frac{4\alpha}{4\alpha+1}} n^{c'c^* - 1}$$

Now choose c^* small enough to guarantee

$-\frac{4\alpha}{4\alpha+1} + c'c^* < -\frac{4\alpha'}{4\alpha'+1}$ i.e. $c'c^* < \frac{4\alpha}{4\alpha+1} - \frac{4\alpha'}{4\alpha'+1}$ which can be done since $\alpha > \alpha'$. ∎