Lecture 12

In this lecture we will use the general scheme of risk estimation to produce confidence ~~set~~ sets to solve the confidence ball problem in the density model.

Set up: $X_1, \ldots, X_n \overset{iid}{\sim} P$ on $([0,1], \mathcal{B}[0,1])$ with $\frac{dP}{d\lambda} = \eta$ (λ being the Lebesgue measure on $[0,1]$). Let $0 < \beta_{min} < \beta_{max} < \infty$ be given along with $M, B > 0$. We want to produce $(1-\alpha)$ Honest Adaptive confidence sets over $\bigcup_{\beta \in [\beta_{min}, \beta_{max}]} \mathcal{P}(\beta, M, B) = \mathcal{P}(\beta_{min}, M, B)$

where $\mathcal{P}(\beta, M, B) = \Big\{ f : [0,1] \to \mathbb{R} : 0 \le f \le B \text{ a.e. } \lambda, \int f \, d\lambda = 1, \sup_{x,y} \frac{|f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(y)|}{|x-y|^{\beta - \lfloor\beta\rfloor}} \le M \Big\}$ where $\lfloor\beta\rfloor$ is the largest integer strictly smaller than β.

We will ~~solve~~ try to understand the problem in two regimes.

(i) Adaptation Regime $(\beta_{max} \le 2\beta_{min})$

(ii) Non-Adaptation Regime $(\beta_{max} > 2\beta_{min})$

As the name suggests, it will be possible to construct honest adaptive confidence sets in case (i) and cannot be done in case (ii).

Let us now focus on these two cases separately.

(i) Adaptation Regime:

We deal with this regime first since we can directly apply the ideas of risk estimation developed in lecture 11 to construct honest adaptive confidence sets.

Theorem: There exists a measurable set $\hat{C}_n \subseteq \left(L_2[0,1], \mathcal{B}(L_2[0,1])\right)$ (based on our data $X_1, \ldots, X_n$, $M, B, \alpha, \beta_{min}, \cancel{\beta_{max}}$) such that one has the following properties hold.

(a) $$\liminf_{n\to\infty} \inf_{\eta \in \mathcal{P}(\beta_{min}, M, B)} \mathbb{P}_\eta\left(\eta \in \hat{C}_n\right) \geq 1-\alpha.$$

(b) For any $\beta \in [\beta_{min}, 2\beta_{min}]$,

$$\limsup_{n\to\infty} \sup_{\eta \in \mathcal{P}(\beta, M, B)} \frac{\mathbb{E}_\eta\left(\mathrm{diam}(\hat{C}_n)\right)}{n^{-\beta/2\beta+1}} \leq C(M, B, \alpha, \beta_{min}, \cancel{\beta_{max}})$$

Proof: For now we will do the proof for $2\beta_{min} < 1$. This is not essential, but since we will be using the Haar Wavelet basis, the approximation properties will be readily available for regularity smaller than 1.

Start with an adaptive estimator of $\hat{\eta}$ of η in the following sense:

(I) $\limsup\limits_{n\to\infty} \sup\limits_{\eta\in\mathcal{P}(\beta,M,B)} \mathbb{E}_\eta \|\eta-\hat{\eta}\|_\infty \lesssim \left(n/\log n\right)^{-\beta/2\beta+1} \times c(M,B,\beta_{min})$

(II) $\liminf\limits_{n\to\infty} \inf\limits_{\eta\in\mathcal{P}(\beta,M,B)} \mathbb{P}_\eta\left(\hat{\eta}\in H(\beta,c)\right) \geq 1-\varepsilon$ for large enough $c := c(M,B,\beta_{min})$ for any $\varepsilon>0$

(III) $\limsup\limits_{n\to\infty} \sup\limits_{\eta\in\mathcal{P}(\beta,M,B)} \mathbb{P}_\eta\left(\{0<\hat{\eta}<2B\}^c\right) \leq \varepsilon$ for any $\varepsilon>0$

<u>Remark</u>: Such an estimator can be constructed from a Lepski type method and then through smooth truncation to obtain desired property (III).

For now consider that we are on the event that

$$\left\{ \|\hat{\eta}-\eta\|_\infty \leq \left(n/\log n\right)^{-\beta/2\beta+1} c(M,B,\beta_{min}),\ \hat{\eta}\in H(\beta,c),\ 0<\eta<2B \right\}$$

which can be assumed to have probability larger than $\alpha/2$ for n sufficiently large (depending on α). Also assume w.l.o.g. that we have a sample of size $2n$ $(X_1,\dots,X_{2n})$ which we split in part 1 $(X_1,\dots,X_n)$ and part 2 $(X_{n+1},\dots,X_{2n})$ with part 2 to construct $\hat{\eta}$. We denote expectation with part j of the sample as $\mathbb{E}_{\eta,j}$, $j\in\{1,2\}$, holding the other parts fixed.

The construction of the confidence set will be now done by estimation $\|\eta-\hat\eta\|_2^2$ from the first part of the sample, uniformly well over $\mathcal{P}(\beta_{min}, M, B)$. This we know how to do.

Let $2^j \sim n^{\frac{2}{4\beta_{min}+1}}$ and let as usual let $K_{V_j}(x_1, x_2)$ be the Haar projection kernel onto $V_j = \text{span}\{\psi_{lk},\ l=-1,\dots,j,\ k=0,\dots 2^l-1\vee 0\}$

We can then estimate $\|\eta-\hat\eta\|_2^2$ by

$$\hat\chi_n = \frac{1}{n(n-1)}\sum_{1\le i_1\ne i_2\le n}\sum_{l=-1}^{j}\sum_{k=0}^{2^l-1\vee 0}\left(\psi_{lk}(x_{i_1}) - \langle\psi_{lk},\hat\eta\rangle\right)\left(\psi_{lk}(x_{i_2}) - \langle\psi_{lk},\hat\eta\rangle\right)$$

This once again can be derived from basic principles of a Von-Mises expansion of $\|\eta-\hat\eta\|_2^2$ as a functional in η or simply by ~~to~~ noting

$$\|\eta-\hat\eta\|_2^2 = \underbrace{\sum_{l=-1}^{j}\sum_{k=0}^{2^l-1\vee 0}\langle\psi_{lk},\eta-\hat\eta\rangle^2}_{\text{estimate this part by } \hat\chi_n} + \underbrace{\sum_{l>j}\sum_{k=0}^{2^j-1}\langle\psi_{lk},\eta-\hat\eta\rangle^2}_{\text{small}}$$

Now by standard calculations,

$$\mathbb{E}_{P,1}(\hat\chi_n) = \|\Pi_{V_j}(\eta-\hat\eta)\|_2^2$$

$$\text{Var}_{P,1}(\hat\chi_n) \le c(B)\left[\frac{2^j}{n^2} + \frac{\|\Pi_{V_j}(\eta-\hat\eta)\|_2^2}{n}\right]$$

(Derive this using our standard variance calculations using projection type kernels)

~~Actually $\text{Var}_{P,1}(\hat\chi_n)$~~

Now identify $\Theta = \mathcal{P}(\beta_{min}, M, B)$, $k_n = 2^j$, $\hat{\tau}^2_{n,k,\theta} = c(B)\left(\frac{2^j}{n^2} + \frac{\|\Pi_{V_j}(\eta-\hat{\eta})\|^2}{n}\right)$,
$R_{k_n,n}(\hat{\theta}^{(n)}) = \hat{\chi}_n$, $\hat{\theta}^{(n)} = \hat{\eta}$ and $B_{k_n} = \|\Pi_{V_j^\perp}(\eta-\hat{\eta})\|_2^2$.

By our theorem, the set

$$\hat{C}_n = \left\{\eta : \|\eta-\hat{\eta}\|_2 \le \sqrt{z_\alpha\sqrt{c(B)}\left\{\frac{2^j}{n^2} + \frac{\|\Pi_{V_j}(\eta-\hat{\eta})\|_2^2}{n}\right\} + \hat{\chi}_n + 2\|\Pi_{V_j^\perp}(\eta-\hat{\eta})\|_2^2}\right\}$$

is a $(1-\alpha)$-honest confidence set over $\mathcal{P}(\beta_{min}, M, B)$ provided

$$\limsup_{n\to\infty} \sup_{\eta\in\mathcal{P}(\beta_{min},M,B)} \mathbb{P}_\eta\left(\hat{\chi}_n - \|\Pi_{V_j}(\eta-\hat{\eta})\|_2^2 \le -z_\alpha\sqrt{c(B)\left(\frac{2^j}{n^2} + \frac{\|\Pi_{V_j}(\eta-\hat{\eta})\|_2^2}{n}\right)}\right) \le \alpha \quad \ldots (*)$$

and

$$\limsup_{n\to\infty} \sup_{\eta\in\mathcal{P}(\beta_{min},M,B)} \mathbb{P}\left(\left|\hat{\chi}_n - \|\Pi_{V_j}(\eta-\hat{\eta})\|_2^2\right| \ge M_n\sqrt{c(B)\left(\frac{2^j}{n^2} + \frac{\|\Pi_{V_j}(\eta-\hat{\eta})\|_2^2}{n}\right)}\right) = 0 \quad \ldots (**)$$

By our bias variance calculations note that $(*)$ and $(**)$ are true by Chebychev's inequality. This proves honesty by choosing $z_\alpha = 1/\sqrt{\alpha}$.

We now have to bound the confidence set diameter.

For $\eta \in \mathcal{P}(\beta, M, B)$, $\beta \in [\beta_{min}, 2\beta_{min}]$, and M large,

$\mathbb{P}_\eta\left(\text{diam}(\hat{C}_n) \geq M n^{-\frac{\beta}{2\beta+1}}\right)$ needs to be bounded.

$$\le \cancel{\mathbb{P}_\eta\left(\left(\frac{2^j}{n^2}\right)^{\frac{1}{4}} + \left(\frac{\|\eta-\hat{\eta}\|_2^2}{n}\right)^{\frac{1}{4}} + \sqrt{\hat{\chi}_n} + \ldots \geq \ldots n^{-\frac{\beta}{2\beta+1}}\right)}$$

$$\cancel{\frac{\|\eta-\hat{\eta}\|_2^2}{n} = O_P\left(n^{-\frac{2\beta}{2\beta+1}}\right)}$$

Once again by the theorem, the diameter is smaller in order than

$$(*) \quad \left(\frac{2^j}{n^2}\right)^{\frac{1}{4}} + \|\Pi_{V_j^\perp}(\eta-\hat{\eta})\|_2^2 + \|\eta-\hat{\eta}\|_2$$

uniformly over any $\mathcal{P}(\beta, M, B)$, $\beta > \beta_{min}$

Now $\hat{\eta}$ is adaptive estimator of η

$\Rightarrow$ $\eta-\hat{\eta}$ is also β smooth & ~~η~~ $\eta-\hat{\eta} \in H(\beta, C)$ for large C w.h.p.

$$\therefore \; * = O_p\left(\left(\frac{2^j}{n^2}\right)^{\frac{1}{4}} + 2^{-j\beta} + n^{-\frac{\beta}{2\beta+1}}\right)$$

$$= O_p\left(n^{-\frac{2\beta_{min}}{4\beta_{min}+1}} + n^{-\frac{\beta}{2\beta+1}}\right) = O_p\left(n^{-\frac{\beta}{2\beta+1}}\right) \text{ if } \beta \le 2\beta_{min}$$

$\Rightarrow$ One can have honest adaptation over $[\beta_{min}, 2\beta_{min}]$.