STA 3000F Lecture 10

Function class $\mathcal{F}$.

$$G_n(r) := \mathbb{E}\left[\sup_{h \in \mathcal{F}^* \cap B(r)} \frac{1}{n}\sum_i^n g_i h(X_i)\right]$$

where g_i's are iid Gaussian.

Rate for constrained LS:

$$r_n = \frac{G_n(r_n)}{r_n}.$$

- Hölder class. in $\mathbb{R}^d$, w/ smoothness order β

$$\log N(\varepsilon) \lesssim \left(\frac{C}{\varepsilon}\right)^{d/\beta}.$$

- $$G_n(r) \leq \frac{1}{\sqrt{n}}\int_0^r \sqrt{\log N(\varepsilon)}\, d\varepsilon.$$

$\beta > d/2$: $$\int_0^r \left(\frac{C}{\varepsilon}\right)^{\frac{d}{2\beta}} d\varepsilon \asymp C' \cdot r^{1-\frac{d}{2\beta}}$$

Finite when $\beta > d/2$.

Substituting back to $r_n^2 = G_n(r_n)$

$$r_n^2 = \frac{C'}{\sqrt{n}} \cdot r_n^{1-\frac{d}{2\beta}}$$

$$r_n = C^K \cdot n^{-\frac{\beta}{d+2\beta}}$$

(Optimal in the $\beta > d/2$ regime).

when $\beta < d/2$.

$$G_n(r) \lesssim \delta_0 + \frac{1}{\sqrt{n}} \int_{\delta_0}^{r} \sqrt{\log N(\varepsilon)}\, d\varepsilon$$

$$= \delta_0 + \frac{1}{\sqrt{n}} \int_{\delta_0}^{r} \varepsilon^{-\frac{d}{2\beta}}\, d\varepsilon$$

$$\lesssim \delta_0 + \frac{1}{\sqrt{n}}\, \delta_0^{1-\frac{d}{2\beta}}.$$

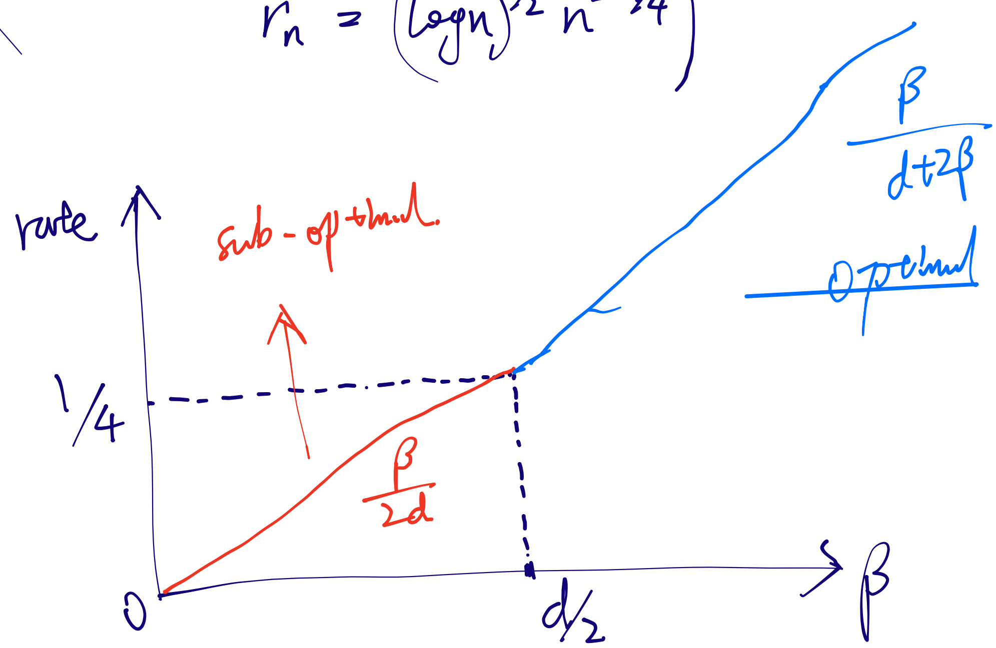
Optimise w.r.t. $\delta_0 = n^{-\beta/d}$

$$G_n(r) \lesssim n^{-\beta/d}$$

$$r_n^2 = G_n(r_n) \Longrightarrow r_n = n^{-\frac{\beta}{2d}}$$

(Critical regime $\beta = d/2$)

$$r_n = (\log n)^{1/2}\, n^{-1/4}$$



— Sub-optimality comes from constrained LS itself

— When $\int_0^{n} \sqrt{\log N(\varepsilon)}\, d\varepsilon = +\infty$

Constrained LS (regularized) is usually sub-optimal.

Projection estimator (sieve).

$$f^* \in \mathcal{F} := S(\beta). \quad := W^{\beta,2}$$

$\varphi_1, \varphi_2, \cdots$ —— real Fourier basis

$$\sum_{j\geq 1}^{+\infty} \langle f^*, \varphi_j \rangle_{L^2}^2 \cdot j^{2\beta} < +\infty.$$

(sometimes, better defn for Sobolev class when β is not integer).

Fixed design setting . $x_i = i/n$.

Discrete Fourier basis .

$$\mathcal{F} = \left\{ f^* : \sum_{j\geq 1}^{n} \langle f^*, \varphi_j \rangle_n^2 \, j^{2\beta} < +\infty \right\}$$

(See Tsybakov for discretization err)

$$\langle f, g\rangle_n := \frac{1}{n}\sum_{i=1}^{n} f(x_i)\, g(x_i)$$

Idea: truncation in frequency domain.

$Y = f^* + \varepsilon$

- Observe $Y \in \mathbb{R}^n$
- $\hat{c}_j := \langle Y, \varphi_j\rangle_n$ (for $j = 1, 2, \dots n$)
- $\hat{f}_n = \sum_{i=1}^{N} \hat{c}_j \varphi_j$.

Analysis:

$$\mathbb{E}\left[\|\hat{f}_n - f^*\|_n^2\right]$$

$$= \sum_{j=1}^{N} \mathbb{E}\left[|\hat{c}_j - \langle f^*, \varphi_j\rangle_n|^2\right] + \sum_{j=N+1}^{n} \langle f^*, \varphi_j\rangle_n^2 .$$

$$= \mathbb{E}\left[\langle \varepsilon, \varphi_j\rangle_n^2\right]$$

$$= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \varphi_j(X_i)\right)^2\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \varphi_j(X_i)^2$$

$$= \frac{1}{n}\|\varphi_j\|_n^2 = \frac{1}{n}$$

$$\mathbb{E}\left[\|\hat{f}_n - f^*\|_n^2\right] \le \frac{N}{n} + \sum_{j=N+1}^{n} \langle f^*, \varphi_j\rangle^2$$

We know $\sum_{j=N+1}^{n} j^{2\beta} \langle f^*, \varphi_j \rangle^2 \leq L^2$

$$\sum_{j=N+1}^{n} \langle f^*, \varphi_j \rangle^2 \leq \frac{1}{(N+1)^{2\beta}} \sum_{j=N+1}^{n} j^{2\beta} \langle f^*, \varphi_j \rangle^2$$

$$\leq \frac{L^2}{N^{2\beta}}$$

Balance $\frac{N}{n} + \frac{1}{N^{2\beta}} \Rightarrow N = n^{\frac{1}{2\beta+1}}$

Rate of convergence $\leq n^{-\frac{2\beta}{2\beta+1}}$.

$(\forall \beta > 0)$. $\mathbb{E}\left[\|\hat{f}_n - f^*\|_n^2\right]$

Density estimation $X_1, X_2, \ldots X_n \overset{iid}{\sim} p \in \mathcal{P}$.

Want to recover p.

Methods 1. MLE. (not covered in class)

$$\hat{p} = \underset{p \in \mathcal{P}}{\operatorname{argmax}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log p(X_i) \right\}$$

(empirical process tools for density class).

2. Local method

Kernel density estimation.

Smooth things out

Setting:

p supported on $[0,1]$

(can be relaxed).

$$\hat{p}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right).$$

. If K is $\frac{1}{2} \mathbb{1}_{[-1,1]}$ local averaging.

. In general, $\int_{\mathbb{R}} K(x)\,dx = 1$. chosen by evaluation.

(So that $\int_{\mathbb{R}} \hat{p}_n(x)\,dx = 1$).

Analysis.

. var.

$$\mathrm{var}\left(\hat{p}_n(x_0)\right) = \frac{1}{nh^2} \mathrm{var}\left(K\left(\frac{X - x_0}{h}\right)\right).$$

$$\leq \frac{1}{nh^2} \int_{\mathbb{R}} K^2\left(\frac{y - x_0}{h}\right) p(y)\,dy$$

$$\leq \frac{p_{max}}{nh} \int_{\mathbb{R}} K^2(x)\,dx.$$

(Assuming $p(x) \leq p_{max}$, $\int_{\mathbb{R}} K^2(x)\,dx < +\infty$).

• Bias

$$\mathbb{E}\left[\hat{p}_n(x_0)\right] - p(x_0)$$

$$= \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{y-x_0}{h}\right)\cdot\left(p(y) - p(x_0)\right) dy$$

(Since $\int K = 1$).

$$= \int_{\mathbb{R}} K(u) \underbrace{\left(p(x_0+uh) - p(x_0)\right)}_{|\cdots| \leq L\cdot(uh)^\beta} du.$$

$(y = x_0 + uh)$.

p : Hölder (β) $(0 < \beta \leq 1)$.

$$|p(x) - p(y)| \leq L\cdot|x-y|^\beta.$$

$$\left|\mathbb{E}\left[\hat{p}_n(x_0)\right] - p(x_0)\right| \leq L\cdot h^\beta \cdot \underbrace{\int |K(u)|\cdot|u|^\beta du}_{\text{Assumed to be fini-}}$$

$$b(x_0)^2 + \sigma(x_0)^2$$

$$\leq c\cdot h^{2\beta} + c'\cdot\frac{1}{nh}$$

Choose $h = n^{\frac{-1}{2\beta+1}}$

$$MSE(x_0) \le C \cdot n^{-\frac{2\beta}{2\beta+1}}. \qquad (0<\beta\le 1).$$

For $\beta > 1$. Taylor expansion.

$$p(x_0+uh) - p(x_0)$$

$$= p'(x_0)\cdot uh + \frac{p''(x_0)}{2}\cdot(uh)^2 + \cdots + \frac{p^{(l-1)}(x_0)}{(l-1)!}\cdot(uh)^{l-1}$$

$$+ \frac{p^{(l)}(x_0+\tau_u uh)}{l!}(uh)^l$$

for some $\tau_u \in [0,1]$.

By careful choice of K.

Substitute to bias term.

$$b(x_0) = \overbrace{\int_{\mathbb{R}} K(u)\, p'(x_0)\, uh\, du + \int_{\mathbb{R}} K(u) \frac{p''(x_0)}{2} (uh)^2 du + \cdots + \int_{\mathbb{R}} \frac{p^{(l-1)}(x_0)}{(l-1)!} K(u)(uh)^{l-1} du}^{=0}$$

$$\int_{\mathbb{R}} \frac{p^{(l)}(x_0)}{l!} K(u)\cdot(uh)^l du = 0$$

$$+ \int_{\mathbb{R}} \frac{p^{(l)}(x_0+\tau uh)}{l!} K(u)\cdot(uh)^l du.$$

l largest integer less than β.

ℓ-th order kernel.

$j = 1, 2, \cdots, \ell$ $\quad \int_{\mathbb{R}} u^j K(u)\, du = 0$

(e.g. $\frac{1}{2}\mathbb{1}_{[-1,1]}$: first order).

(An ℓ-th order needs to be negative somewhere ($\ell \geq 2$)).

Example. Legendre polynomials $(\varphi_k)_{k=0}^{+\infty}$

Orthonormal basis on $L^2([-1,1])$.

$$K(u) := \sum_{m=0}^{\ell} \varphi_m(0) \cdot \varphi_m(u) \cdot \mathbb{1}\{u \in [-1,1]\}$$

- $\int_{\mathbb{R}} K(u)\, du = \sum_{m=0}^{\ell} \varphi_m(0) \int_{-1}^{1} \varphi_m(x)\, dx = 1.$

($1 \leq j \leq \ell$)

- $\int_{\mathbb{R}} K(u) \cdot u^j\, du$ $\quad \left(u^j = \sum_{m=0}^{\ell} b_m \varphi_m\right)$

$$= \int_{-1}^{1} \left(\sum_{m=0}^{\ell} b_m \varphi_m(x)\right) \cdot \left(\sum_{m=0}^{\ell} \varphi_m(0) \varphi_m(x)\right) dx$$

$$= \sum_{m=0}^{\ell} b_m \cdot \varphi_m(0) = u^j \Big|_{u=0} = 0.$$

- $\int_{\mathbb{R}} K(x)^2\, dx < +\infty.$

$$b(x_0) = \int_{\mathbb{R}} \frac{\left|p^{(\ell)}(x_0 + \tau_u uh) - p^{(\ell)}(x_0)\right|}{\ell!} (uh)^{\ell} K(u)\,du.$$

$\leq (\tau_u uh)^{\beta-\ell}$

$$\leq \int_{\mathbb{R}} \frac{L \cdot (|u|h)^{\beta-\ell}}{\ell!} |u|^{\ell} h^{\ell} |K(u)|\,du$$

$$= h^{\beta} \cdot \boxed{\int_{\mathbb{R}} \frac{L \cdot |u|^{\beta} \cdot |K(u)|}{\ell!}\,du}$$

bounded constant.

For $\beta > 1$, we also have

$$MSE(x_0) \leq C \cdot \left(\frac{1}{nh} + h^{2\beta}\right)$$

(Optimize $h_n = n^{-\frac{1}{2\beta+1}}$)

$$= C \cdot n^{-\frac{2\beta}{2\beta+1}}.$$

- Need to use ℓ-th order kernel

$$MISE = \int_{\mathbb{R}} MSE(x)\,dx.$$

$$= \int_{\mathbb{R}} \mathbb{E}\left[\left|\hat{p}_n(x) - p(x)\right|^2\right] dx.$$

Analysis of MISE for Sobolev class

$$\left(\int |p^{(\beta)}(x)|^2 dx \leq L^2 < +\infty\right).$$

$$\cdot \int_{\mathbb{R}} \sigma^2(x)\,dx = \frac{1}{nh^2}\int \operatorname{var}\left(K\left(\frac{X-x}{h}\right)\right)dx$$

$$\leq \frac{1}{nh^2}\int\left(\int K^2\left(\frac{z-x}{h}\right)p(z)\,dz\right)dx$$

$$= \frac{1}{nh}\int K^2(x)\,dx.$$

$$\cdot\ b(x) = \int K(u)\cdot\left(p(x+uh) - p(x)\right)du$$

(using $\beta-1$ order kernel)

$$= \int K(u)\cdot\frac{(uh)^\beta}{\beta!}\int_0^1 (1-\tau)^{\beta-1} p^{(\beta)}(x+\tau uh)\,d\tau\,du.$$

$$\int_{\mathbb{R}} b(x)^2\,dx$$

$$\leq \int\left(\int |K(u)|\cdot|uh|^\beta\int_0^1 \left|p^{(\beta)}(x+\tau uh)\right|d\tau\,du\right)^2 dx.$$

Minkowski inequality

$$\|g(\cdot,u_1)+\cdots+g(\cdot,u_m)\|_{L^2} \leq \|g(\cdot,u_1)\|_{L^2}+\cdots+\|g(\cdot,u_m)\|_{L^2}$$

Generalised Minkowski

$$\left(\int\left|\int g(x,u)\,du\right|^2 dx\right)^{1/2} \leq \int\left(\int g^2(x,u)\,dx\right)^{1/2} du$$

By GM, above term

$$\leq h^{2\beta}\left[\int\left(\int K(u)^2 |u|^{2\beta}\left(\underbrace{\int_0^1 \left|p^{(\beta)}(x+\tau uh)\right| d\tau}_{\text{Cauchy-Schwarz}}\right)^2 dx\right)^{1/2} du\right]^2.$$

$$\leq \int_0^1 \left(p^{(\ell)}(x_0+\tau uh)\right)^2 d\tau$$

$$\leq h^{2\beta}\left(\int |K(u)|\cdot|u|^{\beta}\left(\underbrace{\int_{\mathbb{R}}\int_0^1 \left(p^{(\beta)}(x+\tau uh)\right)^2 d\tau dx}_{\leq L}\right)^{1/2} du\right)^2$$

$$\leq h^{2\beta}\cdot L^2\cdot\left(\int |K(u)|\cdot|u|^{\beta} du\right)^2.$$

$$\text{MISE} \lesssim h^{2\beta}+\frac{1}{nh} \qquad \left(h_n = n^{-\frac{1}{2\beta+1}}\right)$$

$$= n^{-\frac{2\beta}{2\beta+1}}.$$

Lack of asymptotic optimality.

p density on $\mathbb{R}$ $\quad p\in \mathbb{H}^2$. i.e. $\int (p''(x))^2 dx \leq L^2$.

Using 1st order kernel, $\quad \text{MISE} \asymp n^{-4/5}$.

Thm. If we use a second kernel $\quad \|K\|_2^2 < +\infty$.

Then $\forall \varepsilon>0$, take

$$h = n^{-1/5}\varepsilon^{-1}\int K^2(u)\,du$$

We have

$$\limsup_{n\to+\infty} n^{4/5} \cdot \mathbb{E}\int \left(\hat{p}_n(x) - p(x)\right)^2 dx \leq \varepsilon.$$

(By way of contrast, for parametric models, optimal $\hat{\theta}_n$ satisfies $\limsup_{n\to+\infty} n \cdot \mathbb{E}\left[\|\hat{\theta}_n - \theta^*\|^2\right] = \operatorname{Tr}\left(I(\theta^*)^{-1}\right)$.)

"Every density in $\mathbb{H}^2$ is asymptotically smoother than average".

Proof — Claim. $(n\to+\infty,\ nh\to+\infty,\ h\to 0)$.

(i) $\displaystyle\int \sigma^2(x)\,dx = \frac{1}{nh}\int K(u)^2\,du + o\left(\frac{1}{nh}\right)$

(ii) $\displaystyle\int b^2(x)\,dx = \cancel{\frac{h^4}{4}\underbrace{\left(\int u^2 K(u)\,du\right)^2}_{=0 \text{ for second order kernel.}}\cdot\left(\int (p''(x))^2\,dx\right)} + o(h^4).$