STA 3000F lec 21.

$f^* \in \Sigma(\beta)$ $\beta \in (0,1]$. NW achieves $n^{-\frac{2\beta}{2\beta+1}}$

Local poly: $$\hat{f}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_1^n K\left(\frac{X_i - x}{h}\right)} \quad (NW).$$

$$\overset{(\text{Exercise})}{=} \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \theta)^2 K\left(\frac{X_i - x}{h}\right).$$

l-th order poly

$$f(y) \approx f(x) + f'(x)\cdot(y-x) + \frac{f''}{2}(x)(y-x)^2 + \cdots + \frac{f^{(l)}}{l!}(y-x)^l.$$

$$U(u) = \begin{bmatrix} 1 \\ u \\ u^2/2 \\ \vdots \\ u^l/l! \end{bmatrix} \in \mathbb{R}^{l+1} \qquad \theta = \begin{bmatrix} f(x) \\ f'(x)\cdot h \\ \vdots \\ f^{(l)}(x)\cdot h^l \end{bmatrix} \in \mathbb{R}^d.$$

$$\hat{\theta}_n(x) := \underset{\theta \in \mathbb{R}^{l+1}}{\operatorname{argmin}} \sum_1^n \left(Y_i - \theta^T U\left(\frac{X_i - x}{h}\right)\right)^2 K\left(\frac{X_i - x}{h}\right). \quad (*)$$

$$\hat{f}_n(x) := e_1^T \hat{\theta}_n(x).$$

$$\hat{\theta}_n(x) = \underbrace{\left(\frac{1}{nh}\sum_{i=1}^n U\left(\frac{X_i - x}{h}\right) U\left(\frac{X_i - x}{h}\right)^T K\left(\frac{X_i - x}{h}\right)\right)}_{B_{n,x} \in \mathbb{R}^{(l+1)\times(l+1)}}^{-1} \cdot \left[\frac{1}{nh}\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) \cdot U\left(\frac{X_i - x}{h}\right)\right]$$

$$\hat{f}_n(x) = \sum_{i=1}^{n} W_{n,i}(x)\cdot Y_i$$

Key property: For any degree-ℓ poly Q.

$$Q(x) = \sum_{i=1}^{n} W_{n,i}(x)\cdot Q(x_i).$$

Proof:

$$Q(x_i) = \begin{bmatrix} Q(x) \\ Q'(x)\,h \\ \vdots \\ Q^{(\ell)}(x)\,h^{\ell} \end{bmatrix} \cdot U\left(\frac{x_i - x}{h}\right).$$

$\rightarrow \theta$. exactly minimizes $(*)$.

Consequently,

$$\sum_{i=1}^{n} W_{n,i}(x) = 1, \qquad \sum_{i=1}^{n} (x_i - x)^k W_{n,i}(x) = 0$$

for $k = 1, 2, \dots, \ell$.

Analysis of loc poly.

$$b(x_0) = \sum_{i=1}^{n} W_{n,i}(x_0)\left(f^*(x_i) - f^*(x_0)\right)$$

$$= \sum_{i=1}^{n} W_{n,i}(x_0)\cdot\left[\sum_{k=1}^{\ell-1} \frac{(f^*)^{(k)}(x_0)\cdot(x_i - x_0)^k}{k!} + \frac{f^{(\ell)}(x_0 + \tau_i(x_i - x_0))}{\ell!}\cdot (x_i - x_0)^{\ell}\right]$$

$$\leq \sum_{i=1}^{n} \left|W_{n,i}(x_0)\right| \cdot \left|\frac{(f^*)^{(\ell)}(x_0 + \tau_i(x_i - x_0)) - (f^*)^{(\ell)}(x_0)}{\ell!}\right| \cdot \left|x_i - x_0\right|^{\ell}$$

$\leq |x_i - x_0|^{\beta - \ell}$

$$\boxed{\begin{array}{l}\text{Assume } K \text{ supported on } [-1,1] \\ \Longrightarrow W_{n,i}(x_0) = 0 \text{ if } |x_i - x_0| > h.\end{array}}$$

$$|b(x_0)| \leq \sum_{i=1}^{n} |W_{n,i}(x_0)| \cdot h^{\beta}.$$

$$\sigma^2(x_0) = \sum_{i=1}^{n} W_{n,i}^2(x_0) \leq \max_i |W_{n,i}(x_0)| \cdot \sum_{i=1}^{n} |W_{n,i}(x_0)|.$$

Need to bound: (i) $\sum_{i=1}^{n} |W_{n,i}(x_0)|$

(ii) $\max_{1 \leq i \leq n} |W_{n,i}(x_0)|$.

$$|W_{n,i}(x)| = \frac{1}{nh} \cdot \left| e_1^T B_{n,x}^{-1} u\left(\frac{x_i - x}{h}\right) \cdot K\left(\frac{x_i - x}{h}\right) \right|$$

$$\leq \frac{K_{max}}{nh} \cdot |||B_{n,x}^{-1}|||_{op} \cdot \left\| u\left(\frac{x_i - x}{h}\right) \right\|_2 \quad (\text{for } |x_i - x| \leq h)$$

$$\left\| u\left(\frac{x_i - x}{h}\right) \right\|_2 \leq \sqrt{1 + 1 + \frac{1}{(2!)^2} + \cdots + \frac{1}{(\ell!)^2}} \leq 2.$$

$$|W_{n,i}(x)| \leq \frac{2K_{max}}{nh\lambda_0} \quad \text{assuming } \boxed{B_{n,x} \succcurlyeq \lambda_0 I_{\ell+1}}$$

Achievable for

- Equispaced design
- Random design.

$$\sum_{i=1}^{n} |W_{n,i}^*(x)|$$

$$\leq \frac{2K_{max}}{nh\lambda_0} \cdot \boxed{\left| \{ i : |x_i - x| \leq h \} \right|}$$

$$\leq \frac{4K_{max} \cdot a_0}{\lambda_0}$$

Assume: for interval A.

$$\frac{1}{n} \left| \{ i : x_i \in A \} \right| \leq a_0 \cdot \max\{ \text{Leb}(A), \tfrac{1}{n} \}.$$

Putting them together

$$MSE(x_0) \leq b^2(x_0) + \sigma^2(x_0)$$
$$\leq C \cdot h^{2\beta} + \frac{C'}{nh}.$$

$$h_n^* = n^{-\frac{1}{2\beta+1}} \implies MSE \lesssim n^{-\frac{2\beta}{2\beta+1}}.$$

$\forall\, \beta > 0.$

Minimax optimality. $\quad Y_i = f^*(x_i) + \varepsilon_i$

Goal: $\underbrace{\inf_{\hat{T}} \sup_{f^* \in \Sigma(\beta)} \mathbb{E}\left[|\hat{T} - f^*(x_0)|^2 \right]}$ $\gtrsim\ ?$ $\left(n^{-\frac{2\beta}{2\beta+1}} \right).$

$$\geq \inf_{\hat{T}} \sup_{f \in \{f_0, f_1\}} \mathbb{E}\left[|\hat{T} - f^*(x_0)|^2 \right]$$

$\begin{pmatrix}\text{From}\\ \text{Hw1}\end{pmatrix}.$ $\quad \geq \dfrac{(f_0(x_0) - f_1(x_0))^2}{8} \left\{ 1 - d_{TV}(P_0, P_1) \right\}$

P_i : joint distr of $(Y_1, Y_2, \cdots Y_n)$ under $f^* = f_i$.

How to bound d_{TV}?

f : convex w/ $f(1)=0$ $\quad D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right]$.

$f(x) = \frac{1}{2}|x-1|$. $\Rightarrow d_{TV}$

$f(x) = x\log x \Rightarrow D_{KL}$

$f(x) = (x-1)^2 \Rightarrow \chi^2$

$f(x) = (\sqrt{x}-1)^2 \Rightarrow H^2$

Useful facts: $\quad d_{TV}(P,Q) \le \sqrt{\frac{1}{2} D_{KL}(P\|Q)}$.

$$D_{KL}(P\|Q) \le \chi^2(P\|Q).$$

$$P = \bigotimes_{i=1}^n P_i, \quad Q = \bigotimes_{i=1}^n Q_i$$

Tensorization property:

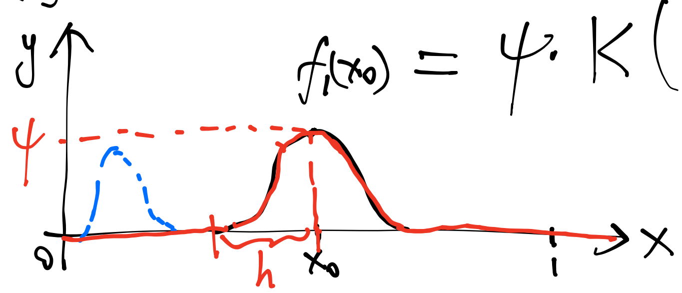
$$D_{KL}(P\|Q) = \sum_{i=1}^n D_{KL}(P_i\|Q_i).$$

$$\chi^2(P\|Q) = \prod_{i=1}^n \left(1+\chi^2(P_i\|Q_i)\right) - 1.$$

Application to nonpar regression:

$f_0(x) = 0 \quad \forall x \in [0,1]$.

$$f_1(x_0) = \psi \cdot K\left(\frac{x-x_0}{h}\right).$$



K supported on $[-1,1]$. $K(0)=1$.

$$K(u) := \exp\left(\frac{-1}{1-u^2}\right)\cdot \mathbb{1}_{\{|u|\le 1\}} \quad \text{"mollifier".}$$

Construct f_1 (choose h, ψ)

$$\begin{cases} f_1 \in \Sigma(\beta) & (1) \\ D_{KL}(P_1\|P_0) \le ?\ \tfrac{1}{2} & (2) \\ |f_1(x_0)-f_0(x_0)| = \psi \text{ as large as possible.} & (*) \end{cases}$$

(1) ℓ: largest int smaller than β

$$f_1^{(\ell)}(x) = \frac{\psi}{h^\ell} K^{(\ell)}\left(\frac{x-x_0}{h}\right).$$

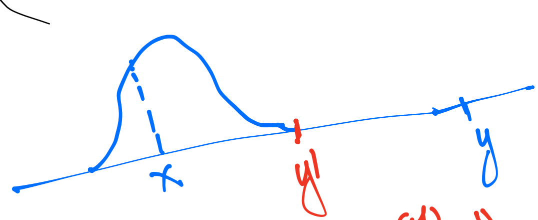
$$\left|f_1^{(\ell)}(x) - f_1^{(\ell)}(y)\right| \le \frac{\psi}{h^\ell}\cdot\left|K^{(\ell)}\left(\frac{x-x_0}{h}\right) - K^{(\ell)}\left(\frac{y-x_0}{h}\right)\right|$$

$$\le C_\ell\cdot\frac{\psi}{h^\ell}\cdot\frac{|x-y|}{h}.$$

$(|x-y|\le 2h)$.

(Assume WLOG $|x-y|\le 2h$)

$$\le 2C_\ell\cdot\frac{\psi}{h^\ell}\cdot\frac{|x-y|^{\beta-\ell}}{h^{\beta-\ell}}$$



$f_1^{(\ell)}(y) = f_1^{(\ell)}(y')$.

Take $\psi = \frac{1}{2C_\ell}\cdot h^\beta$

$$\Longrightarrow f_1 \in \Sigma(\beta).$$

$$\frac{|x-y|}{h} = \left(\frac{|x-y|}{h}\right)^{\beta-\ell}\cdot\underbrace{\left(\frac{|x-y|}{h}\right)^{1-(\beta-\ell)}}_{}\ \le 2.$$

(2). $D_{KL}(P_1 \| P_0) = \sum_{i=1}^{n} D_{KL}(P_{1,i} \| P_{0,i})$ (Y_i indp)

$\boxed{D_{KL}(N(\mu,1) \| N(0,1)) = \mu^2/2.}$

$$= \frac{1}{2} \sum_{i=1}^{n} \left(f_1(X_i) - f_0(X_i)\right)^2.$$

$$\leq \frac{\varphi^2}{2} \cdot \underbrace{\left|\{i : |X_i - x_0| \leq h\}\right|}_{\text{Assumption in local poly}}$$

$$\leq \frac{\varphi^2 a_0}{2} \cdot nh.$$

Put them together,

$$\inf_{\hat{T}} \sup_{f^* \in \Sigma(\beta)} \mathbb{E}\left[|\hat{T} - f^*(x_0)|^2\right] \geq \frac{1}{8} \cdot \varphi^2 \cdot \{1 - \underbrace{\tfrac{1}{2}\varphi\sqrt{a_0 n h}}_{\leq \frac{1}{2}.}\}$$

where $\varphi = C \cdot h^\beta$,

$$h_n = C' n^{-\frac{1}{2\beta+1}}, \quad \varphi_n = C'' \cdot n^{-\frac{\beta}{2\beta+1}}$$

$$\Updownarrow$$

$$\sqrt{a_0} C \cdot h^{\beta+\frac{1}{2}} \cdot \sqrt{n} \leq 1.$$

$$\Updownarrow$$

$$h \leq \left(\frac{1}{n \cdot a_0 \cdot C^2}\right)^{\frac{1}{2\beta+1}}.$$

MISE lower bound?

Two-point doesn't work. (f_0, f_1).

$$\inf_{\hat f} \sup_{f \in \{f_0, f_1\}} \mathbb{E}\left[\|\hat f - f\|_{L^2}^2\right] \geq \frac{\|f_0 - f_1\|_{L^2}^2}{8} \cdot \{1 - d_{TV}(P_0, P_1)\}$$

$$d_{TV}(P_0, P_1) \leq \sqrt{\tfrac{1}{2} D_{KL}(P_0 \| P_1)} = \sqrt{\tfrac{n}{2}} \cdot \|f_0 - f_1\|_n.$$

$\|f_0 - f_1\|_n \leq \frac{c}{\sqrt{n}}$. Cannot hope to get anything better than $O(n^{-1})$.

Recall: Le Cam two-point

Test $H_0: X \sim P_0$ vs. $H_1: X \sim P_1$

"Multiple hypothesis":

$H_1: X \sim P_1$, $H_2: X \sim P_2$, ..., $H_M: X \sim P_M$.

Prior distribution: $J \sim \mathrm{Unif}(\{1, 2, \dots, M\})$.

Data: $X \sim P_J$

Goal: estimate J.

Claim: For any T,

$$\mathbb{P}(T(X) \neq J) \geq 1 - \frac{I(X; J) + \log 2}{\log M}.$$

(Fano's inequality).

$$I(X;J) := D_{KL}\left(\mathbb{P}_{X,J} \,\|\, \mathbb{P}_X \times \mathbb{P}_J\right).$$

For $J \sim \mathrm{Unif}$, $\quad I(X;J) = \frac{1}{M}\sum_{j=1}^{M} D_{KL}(P_j \| \bar{P})$!

$$\bar{P} := \frac{1}{M}\sum_{j=1}^{M} P_j.$$

Detour: Mutual Information.