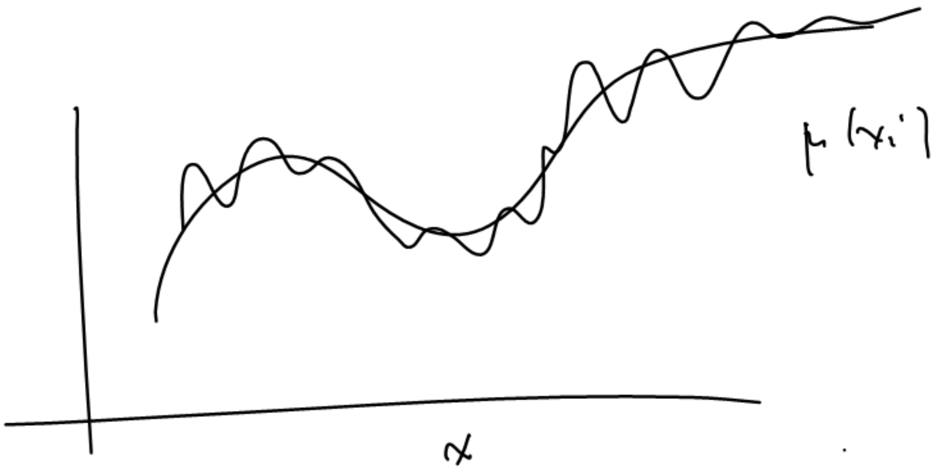
Lecture 23 29 Nov 2016

$$Y_i = \mu(x_i) + \varepsilon_i, \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$



$$\left.\begin{aligned} \mu(x_i) &= \beta_0 + \beta_1 x_i \\ \text{I.}\quad \mu(x_i) &= \sum_{k=0}^{p} \beta_k x_i^k \end{aligned}\right\} \text{parametric models for } \mu(x_i)$$

II. Non parametric representation:

Basis Functions $\{\phi_1(x), \ldots, \phi_Q(x)\}$

$\phi_k(x)$ are orthonormal

$$\sum_i (\phi_k(x_i))^2 = 1$$

$$\langle \phi_k, \phi_{k'} \rangle = \sum_i \phi_k(x_i)\, \phi_{k'}(x_i) = 0$$

$$\mu(x_i) = \sum_{k=1}^{Q} \beta_k \phi_k(x_i)$$

$\{\phi_k\}$: splines, wavelets, Fourier/Trigonometric

etc...

$$\underline{Y} = \underbrace{X \underline{\beta}} + \underline{\varepsilon}$$

$$\underline{Y} = \begin{array}{c} \begin{array}{ccc} \underline{\phi}_1 & \underline{\phi}_2 & \underline{\phi}_Q \end{array} \\ \begin{pmatrix} \phi_1(x_1) & & \phi_Q(x_1) \\ \vdots & \cdots & \vdots \\ \phi_1(x_n) & & \phi_Q(x_n) \end{pmatrix} \end{array} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_Q \end{pmatrix} + \underline{\varepsilon}$$

$$\underline{\hat{\beta}} = (X'X)^{-1} X' \underline{Y}$$

$$X'X = \begin{pmatrix} \underline{\phi}_1' \\ \vdots \\ \underline{\phi}_Q' \end{pmatrix} \begin{pmatrix} \underline{\phi}_1 \cdots \underline{\phi}_Q \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \\ 0 & \cdots & \cdots & 1 \end{pmatrix}$$

$$\underline{\hat{\beta}} = X' \underline{Y}$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_Q \end{pmatrix} = \begin{pmatrix} \underline{\phi}_1' \underline{Y} \\ \vdots \\ \underline{\phi}_Q' \underline{Y} \end{pmatrix}$$

Rem: the estimator for β_j depends only on $\underline{\phi}_j$ and $\underline{Y}$.

Rem. Wavelet estimation typically uses some thresholding:

$$\underline{Y} = X\beta + \varepsilon$$

$\uparrow$ columns of X are wavelets

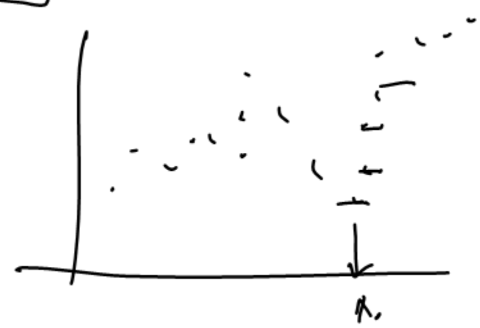
$$\underline{\hat\beta} = X'\underline{Y} \longleftarrow$$

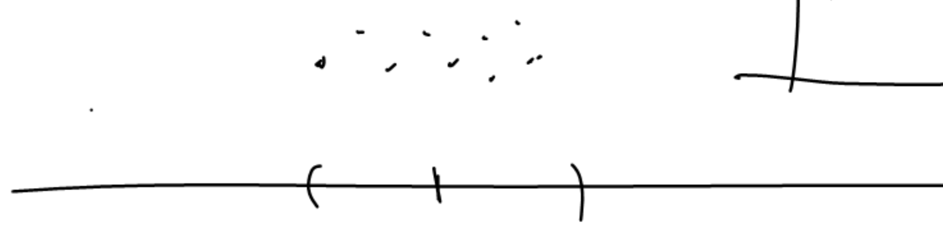
$$\underline{\hat\beta}_{final} = \begin{pmatrix} \hat\beta_{1,final} \\ \vdots \\ \hat\beta_{k,final} \end{pmatrix} \quad \text{where} \quad \text{keep-or-kill}$$

$$\hat\beta_{1,final} = \begin{cases} \hat\beta_1 & \text{if } |\hat\beta_1| > \tau \\ 0 & \text{if } |\hat\beta_1| \leq \tau \end{cases}$$

III. $\mu(x_i)$ "smooth"

Estimate $\mu(x^0)$





x^0

$\cancel{\beta_0 + \beta_1 x^0}$

$$N_r(x^0) = (x^0 - r,\ x^0 + r)$$

$$\mu(x) \approx \mu(x^\circ) \quad \forall\ x \in N_r(x^\circ)$$

$$\Rightarrow \quad \hat{\mu}(x^\circ)$$

$$Y_i \cong \mu(x^\circ) + \varepsilon_i \quad \forall\ x_i \in N_r(x^\circ)$$

$$\hat{\mu}(x_0) = \underset{i:\, x_i \in N_r(x^\circ)}{\text{Ave}} \{ Y_i \}$$

$$\cong \frac{1}{2M+1} \sum_{i:\, x_i \in N_r(x^\circ)} Y_i$$

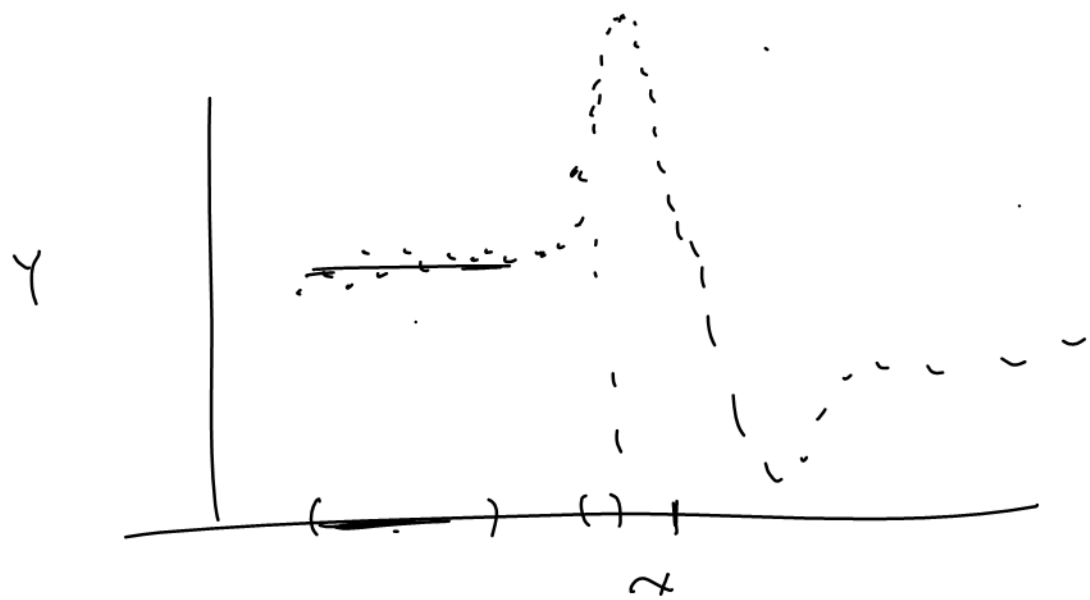
A more general locally weighted estimator

$$\hat{\mu}(x_0) = \sum_i W_i Y_i$$

where $W_i \propto \frac{1}{|x_i - x_0|}$

$$\tilde{W}_i = \exp\left(-\frac{1}{2}(x_i - x_0)^2\right)$$

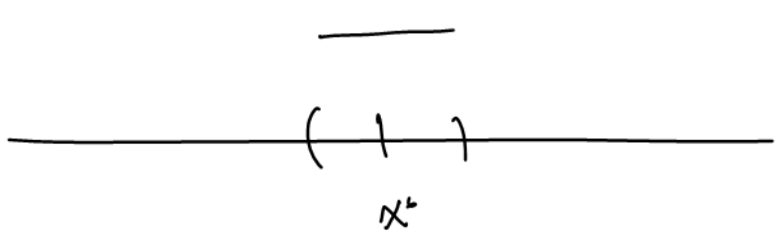
$$i = \frac{\tilde{w}_i}{\sum \tilde{w}_j} \in (0, 1)$$



Recall:

$\mu(x)$ "smooth"

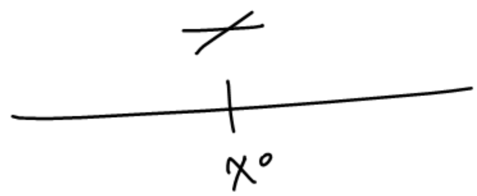
$$\mu(x_i) \cong \mu(x^0) \quad \forall \; x_i \in N_r(x^0)$$



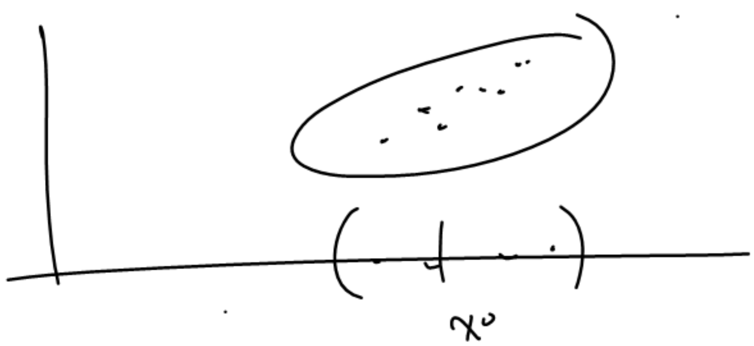
$\mu(x)$ is "locally constant"

$$\mu(x_i) \approx \underline{\underline{\beta_0}}(x^0) \quad \forall\, x_i \in N_r(x^0)$$

$\mu(x)$ is "locally linear"



$$\mu(x_i) \approx \underline{\underline{\beta_0(x^0)}} + \underline{\underline{\beta_1}}(x^0) x_i \quad \forall\, x_i \in N_r(x^0)$$



Define $D(x^0) = \left\{ (x_i, Y_i) : \quad x_i \in N_r(x^0) \right\}$

$$\underline{\hat{\beta}}(x^0) = \begin{pmatrix} \hat{\beta}_0(x^0) \\ \hat{\beta}_1(x^0) \end{pmatrix} = (X^{0\prime} X^0)^{-1} X^{0\prime} \underline{Y}^0$$

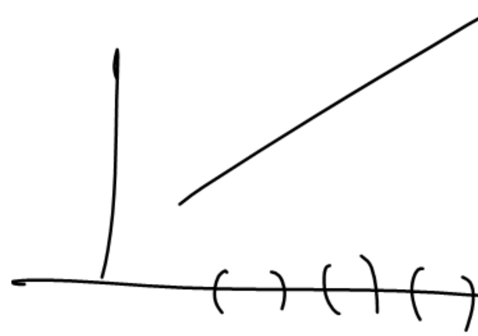
$$\hat{\mu}(x^0) = \hat{\beta}_0(x^0) + \hat{\beta}_1(x^0) \cdot x^0$$

Construct a 95% CI for $\mu(x^0)$

$$\hat{\beta}(x^0) = (X^{0\prime}X^0)^{-1}X^{0\prime}\underline{Y}^0$$

$$\text{cov}(\hat{\beta}(x^0)) = (X^{0\prime}X^0)^{-1} \cdot \cancel{\sigma^2(x^0)} \quad \sigma^2$$

$$\underline{\hat{\beta}} = \underbrace{(X'X)^{-1}X'\underline{Y}}_{\text{global}}$$



$$\underline{\hat{Y}} = X\hat{\beta}$$

$$R_1 = Y_1 - \hat{\mu}(x_1) = Y_1 - (\hat{\beta}_0(x_1) + \hat{\beta}_1(x_1)x_1)$$

$$\vdots \quad Y_2 - \hat{\mu}(x_2) =$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$R_n: \; Y_n - \hat{\mu}(x_n) =$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} R_i^2}{n - df}$$

Hastie & Tibshirani

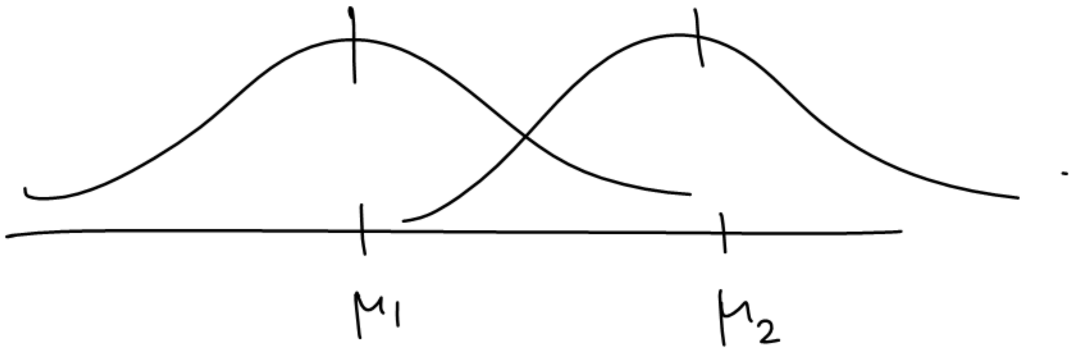
95% CI for $\mu(x^0)$

$$\hat{\beta}_0(x^0) + \hat{\beta}_1(x^0) \cdot x^0 \pm t(.975, n-df) \cdot \sqrt{\underline{c}\,(X^{0\prime} X^0)^{-1}\,\underline{c}'\,\hat{\sigma}^2}$$

$$\mu(x_0) = \underbrace{\overset{\underline{c}}{\begin{pmatrix} 1 & x^0 \end{pmatrix}}} \underbrace{\begin{pmatrix} \beta_0(x_0) \\ \beta_1(x_0) \end{pmatrix}}$$

Computer-intensive / Resampling method of inference

Ex 1. Two indep sample problem



$Y_1^1 \ldots Y_{n_1}^1$ rs from π_1 $N(\mu_1, \sigma^2)$

$Y_1^2 \ldots Y_{n_2}^2$ rs — π_2 $N(\mu_2, \sigma^2)$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \overset{H_0}{\sim} t(df = n_1 + n_2 - 2)$$

reference distn.

if $|t|$ large $\Rightarrow$ reject $H_0: \mu_1 = \mu_2$

Set-up

$Y^1_1 \ldots Y^1_{n_1}$ iid F_{μ_1}

$Y^2_1 \ldots Y^2_{n_2}$ iid F_{μ_2}

Under $H_0: \mu_1 = \mu_2 \Rightarrow$ $Y^1 \ldots$ Y^2 $\rightarrow$

If H_0 true:

$Y^1_1 \ldots Y^1_{n_1} \; Y^2_1 \ldots Y^2_{n_2}$ iid F_μ

Group 1 — Group 2

$U^1_1 , U^1_2 \; U^1_{n_1}$ — $U^2_1 \ldots U^2_{n_2}$

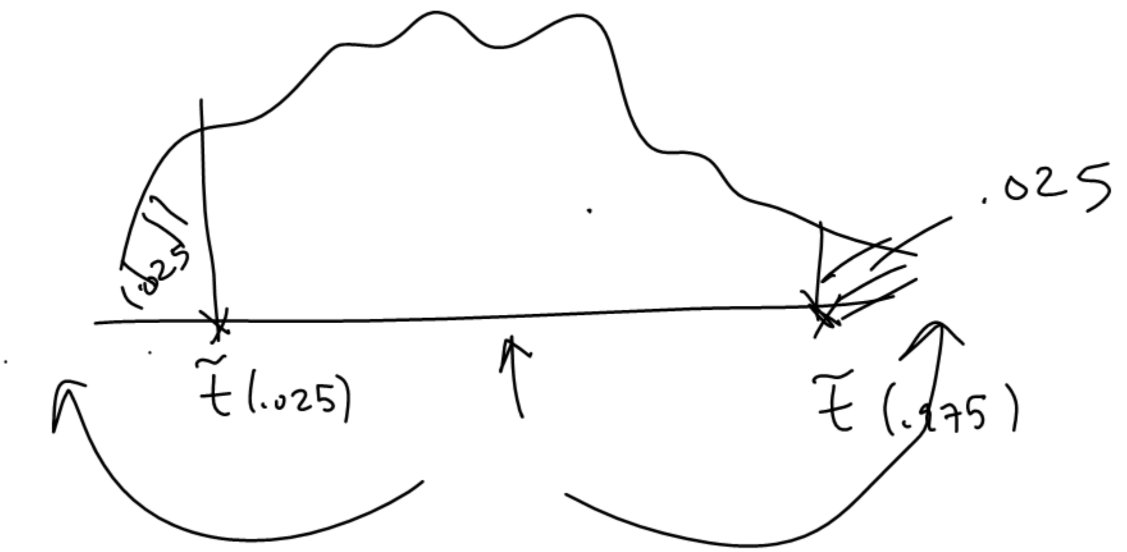
e.g. $Y^1_{10}, Y^2_1, \ldots$

Permutation 1
$$t^{(1)} = \frac{\bar{U}_1^{(1)} - \bar{U}_2^{(1)}}{\sqrt{S_p^{(1)2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$\vdots$

Permutation Q $t^{(Q)}$

Distn $\{t^{(1)} \ldots t^{(q)}\}$



Two Sample Permutation Test

$$(\hat{\beta}_0^1 - \hat{\beta}_0^2)^2 + (\hat{\beta}_1^1 - \hat{\beta}_1^2)^2$$