Last time we saw a good theoretical choice of h.

Also can choose a data dependent method.

$$ISE(h,X) = \int (\hat{f} - f)^2 dt$$

and want

$$h^* = \underset{h>0}{\operatorname{argmin}} \; ISE(h,X)$$

$$ISE = \underbrace{\int \hat{f}^2 dt - 2\int \hat{f} f \, dt}_{J(h)} + \int f^2 dt$$

$$J(h) = R(\hat{f}) - 2\, \mathbb{E}_f \left(\hat{f}_{h,X}(T) \right)$$

$T_n \acute{f}$

$$\overset{(mc)}{\approx} R(\hat{f}) - 2\,\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{n,x}(x_i)$$

$$\overset{(cv)}{\approx} R(\hat{f}) - 2\,\frac{1}{n}\sum_{i=1}^{n} f_{n,x_{[-i]}}(x_i)$$

(Unbiased Cross-Validation)
"UCV"

To construct an approximate confidence interval for our density estimator we can use a smooth version

$$\overline{f_h}(t) = \mathbb{E}_X\left[\hat{f}_{n,x}(t)\right]$$

For a density on (a,b)

$$\ell_n(t) = \hat{f}_n(t) - z\, SE_n(t)$$

$$u_n(t) = \hat{f}_n(t) + z\, SE_n(t)$$

where

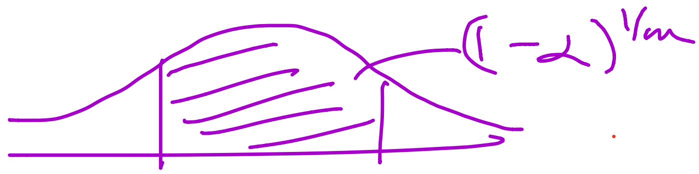
$$SE_n(t) = \frac{S_{\hat{f}}(t)}{n}$$

$$S_{\hat{f}}(t)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i(t) - \overline{Y(t)}\right)^2$$

$$Y_i(t) = \frac{1}{n} K\left(\frac{t - x_i}{n}\right)$$

$$\overline{Y(t)} = \frac{1}{n}\sum_{i=1}^{n} Y_i(t) - \hat{f}_n(t)$$

and



$$z = \Phi^{-1}\left(\frac{1 + (1-\alpha)^{1/m}}{2}\right)$$

$$m = \frac{b-a}{w \cdot h}$$

width of kernel

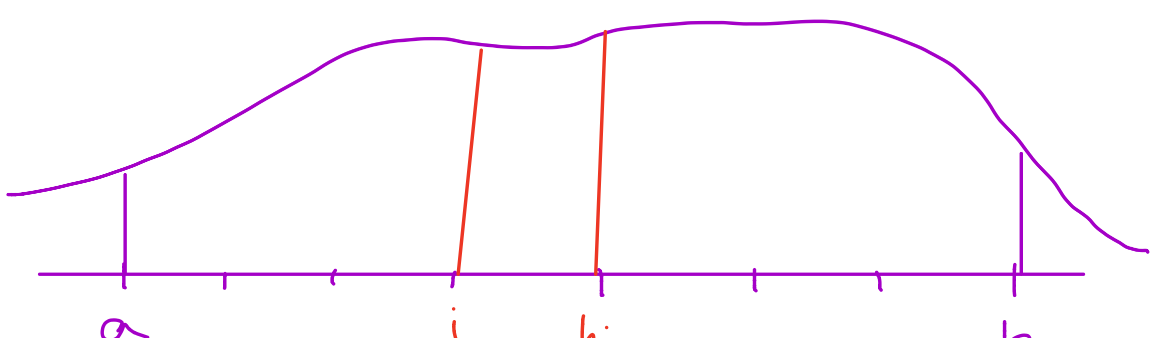
$w=3$ for normal case

$$P(\ell_n(t) \le \bar{f} \le u_n, \; t \in (a,b)] = 1-\alpha$$

$=$

$$\mathbb{P}\left[\hat{f} - z\,SE(t) \le \bar{f} \le \hat{f} + z\,SE(t) : t \in (a,b)\right] = 1-\alpha$$

$$= \mathbb{P}\left[\frac{|\hat{f} - \bar{f}|}{SE_n(t)} \le z, \; t \in [a,b)\right] = 1-\alpha$$



Assume that

$$\frac{|\hat{f}(t) - \bar{f}|}{SE_n(t)} \approx \frac{|f_n(t_j) - \bar{f}_n(t_j)|}{SE_n(t_j)}$$

then

$$\mathbb{P}\left[\frac{|\hat{f} - \bar{f}|}{SE} \le q\right] \approx \prod_{j=1}^{m} \mathbb{P}\left[\frac{|\hat{f}(t_j) - \bar{f}(t_j)|}{SE_n(t_j)} \le q\right]$$

if $\approx$ Z then

$$\approx \prod_{j=1}^{m} \mathbb{P}\left[|Z_j| \le q\right]$$

$$= \prod_{j=1}^{m} (1-\alpha)^{1/m} = 1 - \alpha$$