2.1.6 → Kernel Density Estimation

→ overview → old strategy is to assume pdf form + estimate parameters
→ Another strategy is to estimate the pdf directly w/ Kernel density estimation

→ Def → Kernel function, $K(\cdot)$, is a pdf w/ 2 parameters
→ i^{th} observed value x_i → mean of Kernel function's distribution
→ bandwidth b

→ properties → symmetric about x_i

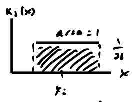
→ Notation → $K_i(\cdot)$ is the kernel function where x_i is the center of its domain
& interpretation → bandwidth b interpretation depends on the $K(\cdot)$ chosen

→ Basic Idea → 1) choose a density form of $K(\cdot)$ → Rectangular, triangular, or Normal
→ 2) Estimate $f(x)$ as the average of $K_1(x), \dots, K_n(x)$

→ Thus, the kernel density estimate of $f(x)$ is

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_i(x) \quad \text{(discrete mixture)}$$

→ Rectangular Kernel →
(uniform)

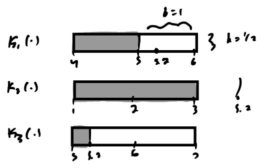


b = half length of rectangle (domain)

→ ex) $x = (5, 2, 6)$, use bandwidth of 1 to 1) estimate $\tilde{f}(5.2)$
2) $P(X < 5.2)$

→ 1) $\tilde{f}(5.2) = \frac{1}{3}[K_1(5.2) + K_2(5.2) + K_3(5.2)]$

sketch kernel centered at each x_i →

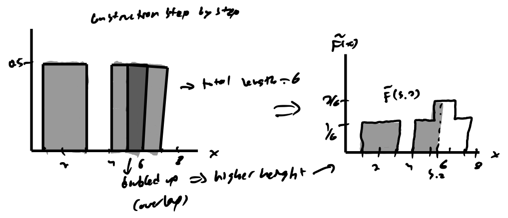


⇒ $\hat{f}(5.2) = \frac{1}{3}(\frac{1}{2} + 0 + \frac{1}{2}) = 1/3$ → $K_i(x) = \begin{cases} \frac{1}{2b} & x_i - b \le x \le x_i + b \\ 0 & \text{otherwise} \end{cases}$

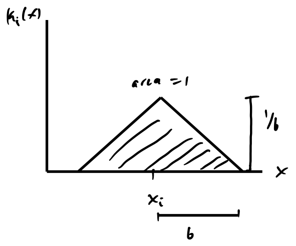
→ 2) $\tilde{F}(5.2) =$  of above ⇒ average

$= \frac{1}{3}\left(\frac{1.2}{2} + 1 + \frac{0.2}{2}\right) = 0.567$

→ $\hat{f}(x)$

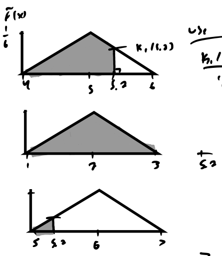


→ Triangular Kernel → assumes isosceles triangle for the kernel function



→ ex) $x = (5, 2, 6)$, use bandwidth of 1 to 1) estimate $\tilde{f}(5.2)$
2) $P(X < 5.2)$

→ 1) $\tilde{f}(5.2) = \frac{1}{3}\{K_1(5.2) + K_2(5.2) + K_3(5.2)\}$
$= \frac{1}{3}[0.8 + 0 + 0.2\} = 1/3$



use ratios or similar triangles → $\frac{\text{small } h}{\text{big } h} = \frac{\text{small } b}{\text{big } b}$

$\frac{K_1(5.2)}{1/b} = \frac{0.8}{b}$, $b=1$ ⇒ $K_1(5.2) = 0.8$

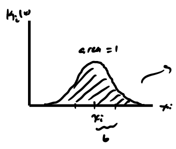
→ $A = \frac{1}{2}bh$

→ 2) $\tilde{F}(5.2) =$ ◺ areas above ⇒ average

$= \frac{1}{3}\left[(0.5 + (0.5 - \frac{0.8(0.8)}{2}) + 1 + \frac{(0.2)(0.2)}{2}\right\}$

$= 0.5667$

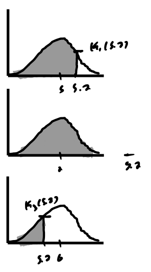
→ Gaussian Kernel → assumes normal density w/ mean x_i + variance b^2 ⇒ sd = b



→ $f(x) = \frac{1}{b\sqrt{2\pi}} \exp\left\{-\frac{1}{2b^2}(x - x_i)^2\right\}$

→ ex) $x = (5, 2, 6)$, use bandwidth of 1 to 1) estimate $\tilde{f}(5.2)$
2) $P(X < 5.2)$

→ 1) $\tilde{f}(5.2) = \frac{1}{3}[0.391 + 0.002 + 0.290] = 0.228$



→ 2) $\tilde{F}(5.2) =$  above ⇒ average

$= \frac{1}{3}\{$ pnorm(5.2, 5, 1) + pnorm(5.2, 2, 1) + pnorm(5.2, 6, 1) $\}$

$= \frac{1}{3}\{0.5793 + 0.9993 + 0.2119\}$

$= 0.5968$