

Kernel Density Estimation

Assume we have $X_1^*, \dots, X_m^*$ a I.I.D. sample from X with density f .

Let $K(x)$ be a Kernel function. That is:

i) $K(x)$ is symmetric

ii) $\int_{-\infty}^{+\infty} K(x) dx = 1$

iii) $\lim_{x \rightarrow +\infty} K(x) = \lim_{x \rightarrow -\infty} K(x) = 0$

$$\text{Let } f_m^*(x) = \frac{1}{m h} \sum_{i=1}^m K\left(\frac{X_i^* - x}{h}\right)$$

Bias. The bias of the KDE is

$$\begin{aligned} E[f_m^*(x)] - f(x) &= E\left[\frac{1}{m h} \sum_{i=1}^m K\left(\frac{X_i - x}{h}\right)\right] - f(x) \\ &= \frac{1}{h} E\left[K\left(\frac{X_i - x}{h}\right)\right] - f(x) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{y - x}{h}\right) f(y) dy - f(x) \end{aligned}$$

Now, we do the change of variable $z = \frac{y - x}{h}$

so that $dz = dy/h$ and we get

$$E[f_m^*(x)] - f(x) = \int_{-\infty}^{+\infty} K\left(\frac{y - x}{h}\right) f(y) dy - f(x)$$

$$= \int_{-\infty}^{+\infty} K(z) f(x + hz) dz - f(x)$$

Next, by a Taylor expansion, when h is small

$$f(x + hz) = f(x) + hz f'(x) + \frac{1}{2} h^2 z^2 f''(x) + o(h^2)$$

where $\lim_{h \rightarrow 0} \frac{o(h^2)}{h^2} = 0$.

Hence,

$$\begin{aligned} E[f_m^*(x)] - f(x) &= \int_{-\infty}^{+\infty} K(z) f(x + hz) dz - f(x) \\ &= \int_{-\infty}^{+\infty} K(z) \left\{ f(x) + hz f'(x) + \frac{1}{2} h^2 z^2 f''(x) + o(h^2) \right\} dz - f(x) \\ &= \int_{-\infty}^{+\infty} K(z) f(x) dz + \int_{-\infty}^{+\infty} K(z) hz f'(x) dz + \int_{-\infty}^{+\infty} K(z) \frac{1}{2} h^2 z^2 f''(x) dz \\ &\quad + o(h^2) - f(x) \\ &= \cancel{f(x)} \underbrace{\int_{-\infty}^{+\infty} K(z) dz}_{"1"} + h f'(x) \underbrace{\int_{-\infty}^{+\infty} K(z) z dz}_{"0"} + \frac{1}{2} h^2 \int_{-\infty}^{+\infty} K(z) z^2 dz + o(h^2) - \cancel{f(x)} \\ &\quad \text{"0" } K \text{ symmetric} \\ &= \frac{1}{2} h^2 \underbrace{\int_{-\infty}^{+\infty} z^2 K(z) dz}_{"1"} + o(h^2) \\ &\quad \text{"1"} \\ &\quad \mu_K \end{aligned}$$

The result in the slides follows, from taking

$$K(x) = \frac{1}{s^* \sqrt{2\pi}} \exp\left(-\left(\frac{x}{s^*}\right)^2\right)$$