2. Let $X_1, \ldots, X_n \sim f$ and let $\widehat{f}_n$ be the kernel density estimator using the boxcar kernel:

$$K(x) = \begin{cases} 1 & -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that

$$\mathbb{E}(\widehat{f}(x)) = \frac{1}{h} \int_{x-(h/2)}^{x+(h/2)} f(y)dy$$

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-X_i}{h}\right)$$

Definition 8.2. The kernel density estimator with bandwidth h and kernel K:

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-X_i}{h}\right)$$

$$\mathbb{E}\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \mathbb{E} K\left(\frac{x-X_i}{h}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \int_{\left|\frac{x-y_i}{h}\right| \le \frac{1}{2}} f(y_i)\, dy_i$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y_i)\, dy_i = \frac{1}{h} \int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)\, dy$$

and

$$\mathbb{V}(\widehat{f}(x)) = \frac{1}{nh^2} \left[\int_{x-(h/2)}^{x+(h/2)} f(y)dy - \left(\int_{x-(h/2)}^{x+(h/2)} f(y)dy \right)^2 \right].$$

$$\mathbb{E}\hat{f}^2(x) = \frac{1}{n^2} \sum_{i,j}^{n} \frac{1}{h^2} K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right)$$

$$= \frac{1}{n^2h^2}\left[\sum_{i=j=1}^{n} \mathbb{E} K\left(\frac{x-X_i}{h}\right) + \sum_{i\neq j} \mathbb{E} K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right)\right]$$

$$= \frac{1}{nh^2} \int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)dy - \frac{n-1}{nh^2} \left(\int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)dy \right)^2$$

$$\Rightarrow \mathrm{Var}(\hat{f}(x)) = \mathbb{E}\hat{f}^2(x) - (\mathbb{E}\hat{f}(x))^2$$

$$= \frac{1}{nh^2}\left[\int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)dy - \left(\int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)dy \right)^2 \right]$$

(b) Show that if $h \to 0$ and $nh \to \infty$ as $n \to \infty$ then $\widehat{f}_n(x) \xrightarrow{\text{P}} f(x)$.

Fix any $\varepsilon > 0$

$P(|X - \mathbb{E}X| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$

$\mathbb{P}\{|\hat{f}_n(x) - f(x)| > \varepsilon\}$

$= \mathbb{P}\{|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) + \mathbb{E}\hat{f}_n(x) - f(x)| > \varepsilon\}$

$\leq \mathbb{P}\{|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| + |\mathbb{E}\hat{f}_n(x) - f(x)| > \varepsilon\}$

$\mathbb{P}\{|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| > \varepsilon\} \leq \frac{1}{\varepsilon^2} Var(\hat{f}_n(x)) \to \frac{1}{\varepsilon^2}\left(\frac{1}{nh} f(x) - \frac{1}{n} f^2(x)\right)$

$\to 0$ as $n \to \infty$, $nh \to \infty$

$\mathbb{E}\hat{f}_n(x) = \frac{1}{h}\int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)\,dy \xrightarrow{h \to 0} \frac{d}{dx}\int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} f(y)\,dy = f(x)$

$\mathbb{P}\{|\hat{f}_n(x) - f(x)| < \varepsilon\} \to 0 \Rightarrow \hat{f}_n(x) \xrightarrow{P} f(x)$

4. Prove equation 6.35.

$$\widehat{J}(h) = \frac{1}{hn^2}\sum_i \sum_j K^*\left(\frac{X_i - X_j}{h}\right) + \frac{2}{nh}K(0) + O\left(\frac{1}{n^2}\right) \qquad (6.35)$$

where $K^*(x) = K^{(2)}(x) - 2K(x)$ *and* $K^{(2)}(z) = \int K(z-y)K(y)dy$.

$\hat{J}(h) = \int \hat{f}^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i)$

$= \frac{1}{n^2h^2}\int \left[\sum_{i=j=1}^{n} K^2\left(\frac{x-X_i}{h}\right) + \sum_{i\neq j}^{n} K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)\right] dx$

$- \frac{2}{n}\sum_{i=1}^{n}\frac{1}{(n-1)h}\sum_{j\neq i} K\left(\frac{x-X_j}{h}\right)$

$= \frac{1}{n^2h}\sum_i\sum_j \int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)dx - \frac{2}{n}\sum_{i=1}^{n}\frac{1}{(n-1)h}\sum_{j\neq i} K\left(\frac{x-X_j}{h}\right)$

By Taylor expansion,

$$= \frac{1}{n^2 h} \sum_i \sum_j K^{(2)}\left(\frac{X_i - X_j}{h}\right) - 2K\left(\frac{X_i - X_j}{h}\right) + \frac{2}{nh} K(0) + O\left(\frac{1}{n^2}\right)$$