Bias and Variance of KDE

- Assume that $X_1, \ldots, X_n$ are IID sample from an unknown density function p.

- In the density estimation problem, the parameter of interest is p, the true density function.

- To simplify the problem, assume that we focus on a given point x0 and we want to analyze the quality of our estimator $\widehat{p}_n(x_0)$

Bias

$$\begin{aligned}
\mathbb{E}(\widehat{p}_n(x_0)) - p(x_0) &= \mathbb{E}\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right)\right) - p(x_0) \\
&= \frac{1}{h}\mathbb{E}\left(K\left(\frac{X_i - x_0}{h}\right)\right) - p(x_0) \\
&= \frac{1}{h}\int K\left(\frac{x - x_0}{h}\right) p(x)dx - p(x_0).
\end{aligned}$$

Change of variable $y = \frac{x - x_0}{h}$

$$\begin{aligned}
\mathbb{E}(\widehat{p}_n(x_0)) - p(x_0) &= \int K\left(\frac{x - x_0}{h}\right) p(x)\frac{dx}{h} - p(x_0) \\
&= \int K\left(y\right) p(x_0 + hy)dy - p(x_0) \qquad \text{(using the fact that } x = x_0 + hy\text{).}
\end{aligned}$$

Bias

Taylor expansion

$$p(x_0 + hy) = p(x_0) - hy \cdot p'(x_0) + \frac{1}{2}h^2y^2p''(x_0) + o(h^2).$$

Plug into the previous expression

$$\begin{aligned}
\mathbb{E}(\widehat{p}_n(x_0)) - p(x_0) &= \int K(y)\, p(x_0 - hy) dy - p(x_0) \\
&= \int K(y) \left[p(x_0) + hy \cdot p'(x_0) + \frac{1}{2}h^2y^2p''(x_0) + o(h^2) \right] dy - p(x_0) \\
&= \int K(y)\, p(x_0) dy + \int K(y)\, hy \cdot p'(x_0) dy + \int K(y)\, \frac{1}{2}h^2y^2p''(x_0) dy + o(h^2) - p(x_0) \\
&= p(x_0) \underbrace{\int K(y)\, dy}_{=1} + hp'(x_0) \underbrace{\int yK(y)\, dy}_{=0} + \frac{1}{2}h^2p''(x_0) \int y^2K(y)\, dy + o(h^2) - p(x_0)
\end{aligned}$$

Bias

Cont'd

$$\begin{aligned}\mathbb{E}(\widehat{p}_n(x_0)) - p(x_0) &= p(x_0) + \frac{1}{2}h^2 p''(x_0)\int y^2 K(y)\,dy - p(x_0) + o(h^2)\\ &= \frac{1}{2}h^2 p''(x_0)\int y^2 K(y)\,dy + o(h^2)\\ &= \frac{1}{2}h^2 p''(x_0)\mu_K + o(h^2),\end{aligned}$$

In conclusion

$$\mathbf{bias}(\widehat{p}_n(x_0)) = \frac{1}{2}h^2 p''(x_0)\mu_K + o(h^2).$$

Observations: 1. $h \to 0$, the bias is shrinking at a rate $O(h^2)$

2. the bias is caused by the curvature (second derivative) of the density function

Variance

$$
\begin{aligned}
\mathsf{Var}(\widehat{p}_n(x_0)) &= \mathsf{Var}\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right)\right) \\
&= \frac{1}{nh^2}\mathsf{Var}\left(K\left(\frac{X_i - x_0}{h}\right)\right) \\
&\leq \frac{1}{nh^2}\mathbb{E}\left(K^2\left(\frac{X_i - x_0}{h}\right)\right) \\
&= \frac{1}{nh^2}\int K^2\left(\frac{x - x_0}{h}\right) p(x)dx \\
&= \frac{1}{nh}\int K^2(y)p(x_0 + hy)dy \quad \text{(using } y = \tfrac{x - x_0}{h} \text{ and } dy = dx/h \text{ again)} \\
&= \frac{1}{nh}\int K^2(y)\left[p(x_0) + hyp'(x_0) + o(h)\right]dy \qquad \text{(Taylor expansion)} \\
&= \frac{1}{nh}\left(p(x_0) \cdot \int K^2(y)dy + o(h)\right) \\
&= \frac{1}{nh}p(x_0)\int K^2(y)dy + o\left(\frac{1}{nh}\right) \\
&= \frac{1}{nh}p(x_0)\sigma_K^2 + o\left(\frac{1}{nh}\right),
\end{aligned}
$$

Observations:

1. variance shrinks at rate O(1/(nh)) when n → ∞ and h → 0.
2. The variance is caused by the density value

Mean Square Error (MSE)

$$\begin{aligned}\mathbf{MSE}(\widehat{p}_n(x_0)) &= \mathbf{bias}^2(\widehat{p}_n(x_0)) + \mathsf{Var}(\widehat{p}_n(x_0)) \\ &= \frac{1}{4}h^4|p''(x_0)|^2\mu_K^2 + \frac{1}{nh}p(x_0)\sigma_K^2 + o(h^4) + o\left(\frac{1}{nh}\right) \\ &= O(h^4) + O\left(\frac{1}{nh}\right).\end{aligned}$$

- Consider the first two terms: AMSE $\frac{1}{4}h^4|p''(x_0)|^2\mu_K^2 + \frac{1}{nh}p(x_0)\sigma_K^2$

$$h_{\mathsf{opt}}(x_0) = \left(\frac{4}{n} \cdot \frac{p(x_0)}{|p''(x_0)|^2}\frac{\sigma_K^2}{\mu_K^2}\right)^{\frac{1}{5}} = C_1 \cdot n^{-\frac{1}{5}}$$

$$\mathbf{MSE}_{\mathsf{opt}}(\widehat{p}_n(x_0)) = O(n^{-\frac{4}{5}})$$

Further reading

http://faculty.washington.edu/yenchic/17Sp_403/Lec7-density.pdf