Lecture #23

- So, if $\hat{f}_h(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-x_i}{h}\right)$, can we describe what happens if the bandwidth h varies?

- Well, recall that K has integral 1: $\int_{\mathbb{R}} K(x)\,dx = 1$. This suggests that K concentrates near the origin. So, let's suppose that in fact K is compactly supported, i.e. $\exists \delta > 0$ s.t. $\int_{-\delta}^{\delta} K(x)\,dx = 1$. In the event K is non-compactly supported, $\forall \varepsilon > 0\ \exists \delta > 0$ s.t. $\int_{-\delta}^{\delta} K(x)\,dx \geq 1-\varepsilon$, so the same is true spiritually, if not literally.

- So, if $\int_{-\delta}^{\delta} K(x)\,dx = 1$, then $\int_{-\delta+x_i}^{\delta+x_i} K(x-x_i)\,dx = 1$

$$\Rightarrow \int_{-[\delta+x_i]\cdot h}^{[\delta+x_i]h} \frac{1}{h} K\left(\frac{x-x_i}{h}\right) dx = 1$$

We conclude that $\frac{1}{h} K\left(\frac{x-x_i}{h}\right)$ concentrates around x_i for h small ($\frac{1}{h} K\left(\frac{x-x_i}{h}\right)$ is tall and skinny), while $\frac{1}{h} K\left(\frac{x-x_i}{h}\right)$ is diffuse for h large ($\frac{1}{h} K\left(\frac{x-x_i}{h}\right)$ is short and fat). You will see this in the Supplemental Question 1 on HW 11.

- As in the case of histogram estimation, the tuneable parameter h can be set to optimally balance between bias and variance:

<u>Theorem (Bias-Variance Tradeoff for KDE)</u>: Let K be a kernel and let $\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-x_i}{h}\right)$ be a kernel density estimate for an underlying density $f(x)$, based on sample $\{x_i\}_{i=1}^{n}$ and bandwidth $h>0$.

(i) The pointwise bias of $\hat{f}_n(x)$ is $B(x) = \mathbb{E}(\hat{f}_n(x)) - f(x)$

$$\approx \frac{1}{2}\sigma_K^2 h^2 \cdot f''(x),$$

where $\sigma_K^2 = \int_{\mathbb{R}} y^2 K(y)\,dx$.

(ii) The pointwise variance of $\hat{f}_n(x)$ is $Var(x) = \mathbb{E}\left(\left[\mathbb{E}(\hat{f}_n(x)) - \hat{f}_n(x)\right]^2\right)$

$$\approx \frac{1}{nh}\left[\int_{\mathbb{R}} K^2(y)\,dy\right]\cdot f(x)$$

(iii) The overall MSE for $\hat{f}_n(x)$ is $MSE = \int_{\mathbb{R}}\left[B(x)^2 + Var(x)\right]dx$

$$\approx \frac{1}{4}\sigma_K^4 \cdot h^4 \cdot \int_{\mathbb{R}}\left[f''(x)\right]^2 dx$$

$$+$$

$$\frac{1}{nh}\int_{\mathbb{R}} K^2(y)\,dy$$

Proof: To see (i), notice $\mathbb{E}(\vec{f}_n(x))$

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

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

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}\frac{1}{h}K\left(\frac{x-x_i}{h}\right)f(x_i)\,dx_i$$

c.o.v.: $y = \frac{x-x_i}{h}$

$dy = -\frac{1}{h}dx_i$

$dx_i = -h\cdot dy$

$x_i = x-hy$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}K(y)f(x-hy)\,dy$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}}K(y)\cdot\left[f(x)-hyf'(x)+\frac{1}{2}[hy]^2f''(x)-\dots\right]dy$$

$$= \frac{1}{n}\sum_{i=1}^{n}\Big[f(x)\cancelto{1}{\int_{\mathbb{R}}K(y)\,dy}-hf'(x)\cancelto{0}{\int_{\mathbb{R}}yK(y)\,dy}$$

$$+\frac{1}{2}hf''(x)\int_{\mathbb{R}}y^2K(y)\,dy - \text{Higher-Order Terms}\Big]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[f(x)+\frac{1}{2}hf''(x)\int_{\mathbb{R}}y^2K(y)\,dy-\dots\right]$$

ignore H.O.T. in h

$$\approx f(x)+\frac{1}{2}hf''(x)\sigma_K^2,$$ as desired. ⌟

⌈To see (ii), we proceed similarly: $\mathrm{Var}(X)$

$$= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-x_i}{h}\right)\right)$$

$$= \frac{1}{n^2}\cdot\frac{1}{h^2}\mathrm{Var}\left(\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)\right)$$

$$= \frac{1}{n}\cdot\frac{1}{h^2}\cdot\mathrm{Var}\left(K\left(\frac{x-x_i}{h}\right)\right)$$

$$= \frac{1}{n}\cdot\frac{1}{h^2}\cdot\mathbb{E}\left(K\left(\frac{x-x_i}{h}\right)^2\right) - \frac{1}{n}\cdot\frac{1}{h^2}\mathbb{E}\left(K\left(\frac{x-x_i}{h}\right)\right)^2$$

$$= \frac{1}{n}\cdot\frac{1}{h^2}\int_{\mathbb{R}} K^2\left(\frac{x-x_i}{h}\right)f(x_i)dx_i - \frac{1}{n}\cdot\frac{1}{h^2}\left[\int_{\mathbb{R}} K\left(\frac{x-x_i}{h}\right)f(x_i)dx_i\right]^2 \quad y=\frac{x-x_i}{h} \text{ as in (i)}$$

$$= \frac{1}{n}\cdot\frac{1}{h^2}\int_{\mathbb{R}} h\cdot K^2(y)\cdot f(x-hy)dy - \frac{1}{n}\cdot\frac{1}{h^2}\left[\int_{\mathbb{R}} h\cdot K(y)\cdot f(x-hy)dy\right]^2$$

$$= \frac{1}{n}\cdot\frac{1}{h}\int_{\mathbb{R}} K^2(y)\cdot f(x-hy)dy - \frac{1}{n}\left[\int_{\mathbb{R}} K(y)f(x-hy)\,dy\right]^2 \Bigg\}\circledast$$

Taylor expanding $f(x-hy) \overset{=}{} f(x) - h[\text{stuff}]$ gives

$$\circledast \overset{=}{} \frac{1}{nh}\int_{\mathbb{R}} K^2(y)\left[f(x)+H.O.T.\right]dy - \frac{1}{n}\int_{\mathbb{R}} K(y)\left[f(x)-H.O.T.\right]dy$$

$$= \frac{1}{nh}\cdot f(x)\cdot\int_{\mathbb{R}} K^2(y)dy + H.O.T. \text{ in } h$$

$$\approx \frac{1}{nh} f(x)\cdot\int_{\mathbb{R}} K^2(y)dy.⌋$$

⌈(iii) is immediate by integrating (i), (ii).⌋

$\textcircled{5}$

- So, if $MSE \approx \frac{1}{4}\sigma_K^4 h^4 \int_{\mathbb{R}} [f''(x)]^2 dx + \frac{1}{nh}\int_{\mathbb{R}} K(x)dx$,

then the optimal h is $h^* = C \cdot n^{-\frac{1}{5}}$ for a constant C depending on f, K but independent of n.

- With such an h^*, $MSE \sim n^{-\frac{4}{5}}$, which goes to 0 as $n \to \infty$ faster than the analogous estimate for histograms, which was $n^{-2/3}$.