To understand bandwidth selection, let us analyze MISE. Suppose that K is a symmetric, continuous probability density function with mean 0 and variance $0 < \sigma_K^2 < \infty$. Let $R(g) = \int g^2(z)dz$. Recall that

$$\text{MISE} = \int \text{MSE}(\hat{f}(x))dx = \int \left[\text{Var}(\hat{f}(x)) + [\text{bias}(\hat{f}(x))]^2 \right] dx$$

Now let $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

Bias: Note $E\{\hat{f}(x)\} = \frac{1}{h}\int K\left(\frac{x-u}{h}\right) f(u)du$

$$= \int K(t) f(x-ht)dt \quad \text{(change of variable)}$$

and using Taylor's expansion, $f(x-ht) = f(x) - ht f'(x) + \frac{h^2t^2}{2} f''(x) + o(h^2)$.

$$\Rightarrow E\{\hat{f}(x)\} = f(x) + \frac{h^2\sigma_K^2}{2} f''(x) + o(h^2)$$ because K is symmetric about 0.

So, $$[\text{bias}\{\hat{f}(x)\}]^2 = \frac{h^4\sigma_K^4}{4}[f''(x)]^2 + o(h^4).$$

$$\Rightarrow \text{ISB} = \int [\text{bias}\{\hat{f}(x)\}]^2 dx = \frac{h^4\sigma_K^2}{4} R(f'') + o(h^4).$$

Variance: $\text{Var}\{\hat{f}(x)\} = \frac{1}{n}\text{Var}\left\{\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right\}$

change of variables (same as above) $$= \frac{1}{nh}\int K(t)^2 f(x-ht)dt - \frac{1}{n}\left[E\left\{\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right\}\right]^2$$

Taylor's expansion (same as above) $$= \frac{1}{nh}\int K(t)^2[f(x)+o(1)]dt - \frac{1}{n}[f(x)+o(1)]^2$$

$$= \frac{1}{nh} f(x) R(K) + o\left(\frac{1}{nh}\right).$$

$$\Rightarrow \text{IV} = \int \text{Var}\,\hat{f}(x)dx = \frac{R(K)}{nh} + o\left(\frac{1}{nh}\right).$$

and $$\text{MISE} = \underbrace{\frac{R(K)}{nh} + \frac{h^4\sigma_K^2}{4}R(f'')}_{\text{AMISE}} + o\left(\frac{1}{nh} + h^4\right).$$

To minimize AMISE with respect to h, seek value of h that avoids excessive bias and variance.

optimal bandwidth $h_0 = \left(\dfrac{R(K)}{n \sigma_K^4 R(f'')} \right)^{1/5}$

$\Rightarrow$ minimal AMISE: $\text{AMISE}_0 = \frac{5}{4} \left[\sigma_K R(K) \right]^{4/5} R(f'')^{1/5} n^{-4/5}$

Recall for histograms, $\text{AMISE}_0 = \left[\dfrac{R(f')}{16} \right]^{1/3} n^{-2/3}$

w/ kernel density estimator, getting closer to parametric of $\frac{1}{n}$.

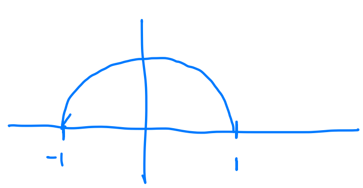
The term $R(f'')$ measures the roughness of the true underlying density. In general, rougher densities are more difficult to estimate and require smaller bandwidth.

The term $[\sigma_K R(K)]^{4/5}$ is a function of the kernel function K.

We could choose K to minimize $[\sigma_K R(K)]^{4/5}$:

If K restricted to be a proper density, minimizer is a scaled version of a quadratic density:

$$K(u) = \frac{3}{4}(1-u)^2 \mathbb{I}(|u| \le 1).$$



"Epanechnikov Kernel", more later.

3.1.1 Cross Validation

We want to evaluate quality of $\hat{f}$ as an estimator of f without using data twice (once for fitting $\hat{f}$ and once for evaluating).

$\Rightarrow$ again use $\hat{f}_{-i}(X_i) = \frac{1}{h(n-1)} \sum_{j \neq i} K\left(\frac{X_i - X_j}{h}\right)$ and let $\hat{Q}(h)$ be a function of $\hat{f}_{-i}(X_i)$ that assesses quality of fit.

e.g. ISE from before.

If $\hat{Q} = ISE$, choose h to minimize $R(\hat{f}) - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i)$.

could instead choose $\hat{Q}(h)$ as as the pseudo likelihood $PL(h) = \prod_{i=1}^{n} \hat{f}_{-i}(X_i)$ and choose h to maximize!

Typically, will be undersmoothed $\Rightarrow$ too bumpy.

3.1.2 Plug-in Methods

If the reference density f is Gaussian and a Gaussian kernel K is used,

$$h_0 = 1.059\, \sigma\, n^{-1/5}$$

sample variance or IQR.

This will probably oversmooth.

Empirical estimation of $R(f'')$ may be a better option.

We could use a kernel density estimator for f'':

$$\hat{f}''(x) = \frac{d^2}{dx^2}\left\{\frac{1}{nh_0}\sum_{i=1}^{n} L\left(\frac{x - x_i}{h_0}\right)\right\}$$

$$= \frac{1}{nh_0^3}\sum_{i=1}^{n} L''\left(\frac{x - x_i}{h_0}\right)$$

where h_0 = bandwidth, L is a sufficiently differentiable kernel to estimate f''.

Note: estimating f and f'' (or $R(f'')$) will require diff'nt bandwidth.

If we use h_0, L to estimate $R(f'')$ and h, K to estimate f, then AMISE of $R(f'')$ is minimized when $h_0 \propto n^{-1/5}$

$$\Rightarrow h_{0,opt} = C_1(R(f''), R(f'''))\, C_2(L)\, h^{5/7}$$

use Gaussian Plugin rule.

$\Rightarrow$ 2 stage solution: (Sheather-Jones method).

1. Use Gaussian plugin rule for h_0 (the bandwidth used to estimate $R(f'')$).
2. h is calculated using $h = \left(\frac{R(K)}{n\sigma_K^4 \underbrace{\hat{R}(f'')}_{\text{comes from } 1}}\right)^{1/5}$ used to produce the final kernel density estimate.

3.2 Choice of Kernel

There are two choices we have to make to perform density estimation:

The kernel and the bandwidth.

The shape of the kernel is much less important.

Recall $AMISE_0 = \frac{5}{4}\left[\sigma_K R(K)\right]^{4/5} R(f'')^{1/5} n^{-4/5}$

3.2.1 Epanechnikov Kernel

The *Epanechnikov kernel* results from choosing K to minimize $[\sigma_K R(K)]^{4/5}$, restricted to be a symmetric density with finite moments and variance equal to 1

$\Rightarrow K(u) = \frac{3}{4}(1-u^2)\mathbb{I}(|u| \le 1)$

and $\sigma_K R(K) = \frac{3}{(5\sqrt{5})}$

$\Rightarrow$ the ratio of $\frac{\sigma_K R(K)}{3/(5\sqrt{5})}$ (*) gives a measure of relative inefficiency of other kernel.

Kernel	Form	Inefficiency ratio (*)
Epanechnikov	$\frac{3}{4}(1-u^2)\mathbb{I}(\lvert u\rvert \le 1)$	1
Biweight	$\frac{15}{16}(1-u^2)^2\mathbb{I}(\lvert u\rvert \le 1)$	1.0061
Triweight	$\frac{35}{32}(1-u^2)^3\mathbb{I}(\lvert u\rvert \le 1)$	1.0135
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$	1.0513
Uniform	$\frac{1}{2}\mathbb{I}(\lvert u\rvert \le 1)$	1.0758

Kernel choice doesn't make much of a difference!

3.2.2 Canonical Kernels

Unfortunately a particular value of h corresponds to a different amount of smoothing depending on which kernel is being used.

Let h_K and h_L denote the bandwidths that minimize AMISE when using symmetric kernel densities K and L. Then,

$$\frac{h_K}{h_L} = \frac{\left(R(K)/\sigma_K^4\right)^{1/5}}{\left(R(L)/\sigma_L^4\right)^{1/5}} = \frac{\delta(K)}{\delta(L)}$$

$\Rightarrow$ the change from bandwidth h for kernel K to a bandwidth that gives the same amount of smoothing for L, use bandwidth $h\,\delta(L)/\delta(K)$.

Epanechnikov: $\delta(K) = 15^{1/5}$
Gaussian: $\delta(K) = \left(1/(2\sqrt{\pi})\right)^{1/5}$
Uniform: $\delta(K) = (9/2)^{1/5}$

Suppose we rescale a kernel shape so that $h = 1$ corresponds to a bandwidth of $\delta(K)$,

The kernel density estimator can then be written as $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h\delta(K)}(x - x_i)$

where $K_{h\delta(K)}(z) = \frac{1}{h\delta(K)} K\left(\frac{z}{h\delta(K)}\right)$.

called a "canonical kernel"

benefit: a single value of h can be used interchangeably for each canonical kernel without affecting the smoothing.

$\Rightarrow$ For a canonical kernel w/ bandwidth h,

$$AMISE(h) = \left(\sigma_K R(K)\right)^{4/5} \left(\underbrace{\frac{1}{nh} + \frac{h^4 R(f'')}{4}}_{\text{bias/variance tradeoff}}\right)$$

So, the Epanechnikov kernel is optimal for any degree of smoothing.

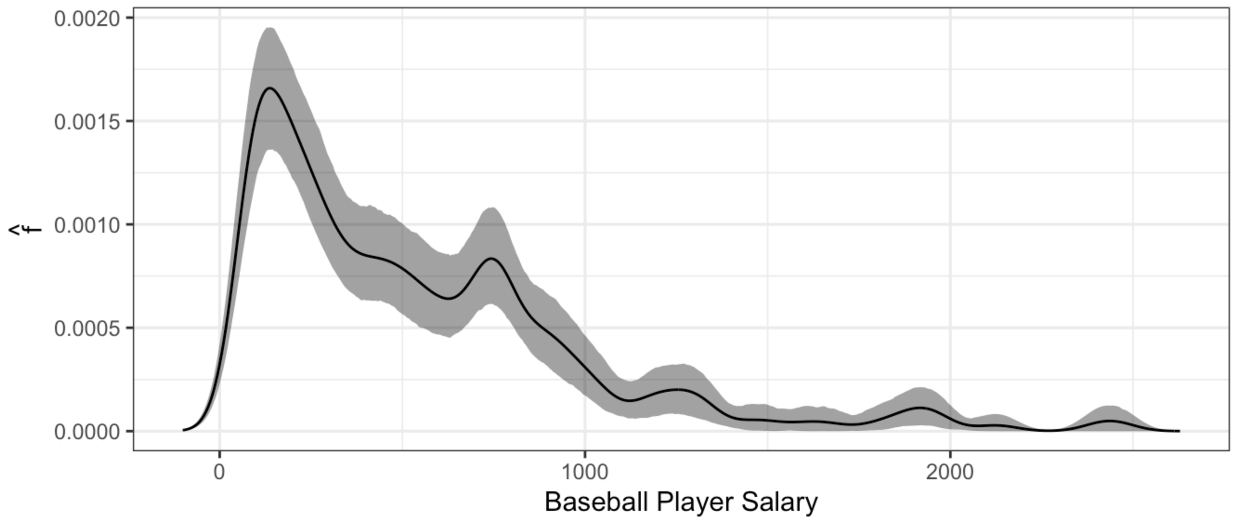
3.3 Bootstrapping and Variability Plot

repeat many times

① get bootstrap sample

② choose bandwith for sample from ① using Sheather-Jones, density estimate obtained

values of $\hat{f}$ recorded at a fixed grid of values.



Note this is NOT a 95% CI

just representation of variability in process of estimating $\hat{f}$.

↳ not telling us $P(\hat{f}(x)_{lower} \le f(x) \le \hat{f}_{upper})$. because bias $\neq 0$.