1. Suppose that $X_1, X_2, \dots, X_n \sim f$. From these, we construct a kernel estimate of $f$, $\widehat{f}$. Now, suppose that $Y \sim \widehat{f}$. We look to investigate $\mathbb{E}(Y)$ and $Var(Y)$.

$$\mathbb{E}(Y) = \int y\widehat{f}(y)dy = \int y\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-X_i}{h}\right)dy = \frac{1}{nh}\sum_{i=1}^{n}\int yK\left(\frac{y-X_i}{h}\right)dy$$

Now completing a substitution $z = \frac{y-X_i}{h}$, we have

$$= \frac{1}{n}\sum_{i=1}^{n}\int (zh + X_i)K(z)dz = \frac{1}{n}\sum_{i=1}^{n}\left[h\int zK(z)dz + X_i\int K(z)dx\right] = \sum_{i=1}^{n}X_i = \overline{X}$$

Where the last equality is due to the fact that $K$ has zero mean and integrates to one. Here, we see that the kernel estimator does not recover the true mean of the data, but recovers the sample mean. Now we compute the variance in a similar fashion.

$$\mathbb{E}(Y^2) = \int y^2\widehat{f}(y)dy = \int y^2\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-X_i}{h}\right)dy = \frac{1}{nh}\sum_{i=1}^{n}\int y^2K\left(\frac{y-X_i}{h}\right)dy$$

Again, we complete a substitution of the form $z = \frac{y-X_i}{h}$ to yield

$$= \frac{1}{n}\sum_{i=1}^{n}\int (X_i + zh)^2K(z)dx = \frac{1}{n}\sum_{i=1}^{n}\left[X_i^2\int K(z)dz + 2X_ih\int zK(z)dz + h^2\int z^2K(z)dz\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 + h^2\mu_2(K)\right) = \frac{1}{n}\sum_{i=1}^{n}X_i^2 + h^2\mu_2(K)$$

Now combining these results we see

$$Var(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{1}{n}\sum_{i=1}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 + h^2\mu_2(K)$$

Here, we see that variance of $Y$ is a function of the variance of the sample as well as the *variability of the kernel*. This feature shows that the variance both scales with the bandwidth and takes into account the variability of the kernel. That is, it does not recover the variability of the population.

2. To calculate Bias of $\widehat{f}''$, we first find the expectation of the of the estimator.

$$\mathbb{E}(\widehat{f}''(x)) = \frac{1}{nh^3}\sum_{i=1}^{n}\mathbb{E}\left[K''\left(\frac{x-X_i}{h}\right)\right] = \frac{1}{h^3}\mathbb{E}\left[K''\left(\frac{x-X_i}{h}\right)\right]$$

where the second equality is because $X_i$ iid $f$ so the argument, regardless of $X_i$ are equivalent in expectation. From here, we complete a substitution for $z = \frac{x-u}{h}$

$$= \frac{1}{h^3}\int K''\left(\frac{x-u}{h}\right)f(u)du = \frac{1}{h^2}\int K''(z)f(x-zh)dz$$

Now, we look to reduce this integral to an integral that contains a $K(\cdot)$. We will achieve this by integrating by parts (twice). Letting, $u = f(x-zh)$, $dv = K''(z)dz$, we have

$$= \frac{1}{h^2}f(x-zh)K'(z)\Big|_{-\infty}^{\infty} + \frac{1}{h}\int K'(z)f'(x-zh)dz = \frac{1}{h}\int K'(z)f'(x-zh)dz$$

We see that the first term goes to zero due to fact that $f$ has finite integral (i.e. tails goes to zero). Now, integrating by parts again with $u = f'(z-zh)$ and $dv = K'(z)dz$, we have

$$= \frac{1}{h}f'(x-zh)K(z)\Big|_{-\infty}^{\infty} + \frac{1}{h}\int hK(z)f''(x-zh)dz = \int K(z)f''(x-zh)dz$$

We see the first terms goes to zero in this case due to the finite integral of $K(\cdot)$ (i.e. tails to zero). Now, we expand $f''(x-zh)$ around the point $x$ to yield

$$\begin{aligned}
&= \int K(z)\Big(f''(x) - zhf^{(3)}(x) + \frac{f^{(4)}(x)z^2h^2}{2} + o(h^2)\Big)dz \\
&= f''(x)\int K(z)dz - hf^{(3)}(x)\int zK(z)dz + \frac{f^{(4)}(x)}{2}\int z^2h^2K(z)dz + o(h^2) \\
&= f''(x) + \frac{f^{(4)}(x)h^2}{2}\mu_2(K) + o(h^2)
\end{aligned}$$

From here, we compute the bias to be

$$Bias(\widehat{f}''(x)) = \mathbb{E}(\widehat{f}''(x)) - f''(x) = \frac{f^{(4)}(x)h^2}{2}\mu_2(K) + o(h^2)$$

Now, to calculate the variance, we have

$$\begin{aligned}
Var(\widehat{f}''(x)) &= \frac{1}{n^2h^6}\sum_{i=1}^{n} Var\left[K''\left(\frac{x-X_i}{h}\right)\right] = \frac{1}{nh^6}Var\left[K''\left(\frac{x-u}{h}\right)\right] \\
&= \frac{1}{nh^6}E\left[K''\left(\frac{x-u}{h}\right)^2\right] - \frac{1}{nh^6}E\left[K''\left(\frac{x-u}{h}\right)\right]^2
\end{aligned}$$

Focusing on the first term, we have via substitution $z = \frac{x-u}{h}$

$$\frac{1}{nh^6}\int K''\left(\frac{x-u}{h}\right)f(u)du = \frac{1}{nh^5}\int K''(z)^2 f(zh+x)dz = \frac{1}{nh^5}\int K''(z)^2\big[f(x) + O(1)\big]dz$$

Here we see that the last equality is due to the Taylor expansion around the point $x$. Finally, this yields

$$\frac{1}{nh^6}E\left[K''\left(\frac{x-u}{h}\right)^2\right] = \frac{f(x)}{nh^5}||K''||_2^2 + o(\frac{1}{nh^5})$$

Now, for the second term, we note that this exactly value was calculated in the bias calculation above. That is

$$\frac{1}{nh^6}E\left[K''\left(\frac{x-u}{h}\right)\right] = \frac{f''(x)}{nh^3} + \frac{f^{(4)}(x)}{2nh^5}\mu_2(K) + o(\frac{1}{nh^3})$$

Now, notice when we square this term, we have a term on the order of $o(\frac{1}{nh^6})$. Plugging this term back into our original variance formula shows

$$Var(\widehat{f}'') = \frac{f(x)}{nh^5}||K''||_2^2 + o(\frac{1}{nh^5}) - o(\frac{1}{n^2h^6}) = \frac{f(x)}{nh^5}||K''||_2^2 + o(\frac{1}{nh^5})$$

Using these facts, we look to minimize MISE as a function of the bandwidth to get an approximate optimal bandwidth value for $\widehat{f}''$.

$$MISE(\widehat{f}'') = \int MSE(\widehat{f}''(x))dx = \int Var(\widehat{f}''(x)) + Bias(\widehat{f}''(x))dx$$

$$= \int \frac{f(x)}{nh^5}||K''||_2^2 + o(\frac{1}{nh^5}) + \left(\frac{h^2}{2}\mu_2(K)f^{(4)}(x) + o(h^2)\right)^2 dx$$

$$= \int \frac{f(x)}{nh^5}||K''||_2^2 + \frac{h^4}{4}\mu_2^2(K)f^{(4)}(x)^2 + o(h^4)dx$$

$$= \frac{||K''||_2^2}{nh^5} + \frac{h^4}{4}\mu_2^2(K)||f^{(4)}||_2^2 + o(h^4)$$

Now differentiating with respect to $h$ and setting to zero we have

$$\frac{-5||K''||_2^2}{nh^6} + 3h^3\mu_2^2(K)||f^{(4)}(x)||_2^2 \overset{set}{=} 0$$

Solving for $h$ gives

$$h_{opt} = \left(\frac{5||K''||_2^2}{n\mu_2^2(K)||f^{(4)}||_2^2}\right)^{1/9}$$

This estimate of the optimal bandwidth follows a very similar structure for that of $\widehat{f}$. Notably, it relies on a derivative of $f$, as well as features of the selected kernel. Moreover, we see that is scales at a rate of $\frac{1}{9} < \frac{1}{5}$. Moreover, we see the effect of the $\widehat{f}$ in that there is an additional constant $\frac{5}{n}$.

3. (a)

$$\int K_2(t)dt = \int cK_1(ct)dt = c\int K_1(z)\frac{dz}{c} = \int K(Z)dz = 1$$

Where the second equality is due to a substitution by $z = ct$.

(b)

$$\int tK_2(t)dt = \int ctK_1(ct)dt = c\int \frac{z}{c}K_1(z)\frac{dz}{c} = \frac{1}{c}\int zK_1(z)dz = 0$$

Again, the second equality is due to a substitution by $z = ct$.

(c)

$$\begin{aligned}
\left(||K_2(t)||_2^2\right)^{4/5}\left(\mu_2(K_2(t))\right)^{2/5} &= \left(||cK_1(ct)||_2^2\right)^{4/5}\left(\mu_2(cK_1(ct))\right)^{2/5} \\
&= \left(\int [cK_1(ct)]^2 dt\right)^{4/5}\left(\int (ct)^2 cK_1(ct)dt\right)^{2/5} \\
&= c^{8/5}\left(\int K_1^2(ct)dt\right)^{4/5} c^{4/5}\left(\int t^2 cK_1(ct)dt\right)^{2/5} \\
&= c^{8/5}\left(\int K_1^2(z)\frac{dz}{c^2}\right)^{4/5} c^{4/5}\left(\int \left(\frac{z}{c}\right)^2 cK_1(z)\frac{dz}{c}\right)^{2/5} \\
&= \frac{c^{8/5}}{c^{8/5}}\left(\int K_1^2(z)dz\right)^{4/5}\frac{c^{4/5}}{c^{4/5}}\left(\int z^2K_1(z)dz\right)^{2/5} \\
&= \left(||K_1(z)||_2^2\right)^2\left(\mu_2(K_1(z))\right)^{2/5}
\end{aligned}$$

The fourth equality is from the substitution $z = ct$ and $z = cu$. Note in the first term, we substitute for a squared function, so we scale by $\frac{dz}{c^2}$ not $\frac{dz}{c}$.

(d)

$$\begin{aligned}
&\left(||K_2(t)||_2^2\right)^{4/5}\left(\mu_2(K_2(t))\right)^{2/5} \\
&= \left(\int \left[c\frac{3}{4}(1-(ct)^2)I(|ct|\le 1)\right]^2 dt\right)^{4/5}\left(\int (ct)^2 c\frac{3}{4}(1-(ct)^2)I(|ct|\le 1)dt\right)^{2/5} \\
&= c^{8/5}\left(\int_{-1/c}^{1/c}\left[\frac{3}{4}(1-(ct)^2)\right]^2 dt\right)^{4/5} c^{4/5}\left(\int_{-1/c}^{1/c} ct^2\frac{3}{4}(1-(ct)^2)dt\right)^{2/5} \\
&= c^{8/5}\left(\int_{-1}^{1}\left[\frac{3}{4}(1-z^2)\right]^2\frac{dz}{c^2}\right)^{4/5} c^{4/5}\left(\int_{-1}^{1} c\left(\frac{z}{c}\right)^2\frac{3}{4}(1-z^2)\frac{dz}{c}\right)^{2/5} \\
&= \frac{c^{8/5}}{c^{8/5}}\left(\int_{-1}^{1}\left[\frac{3}{4}(1-z^2)\right]^2 dz\right)^{4/5}\frac{c^{4/5}}{c^{4/5}}\left(\int_{-1}^{1} z^2\frac{3}{4}(1-z^2)dz\right)^{2/5} \\
&= \left(\int_{\mathbb{R}}\left[\frac{3}{4}(1-z^2)\right]^2 I(|z|\le 1)dz\right)^{4/5}\left(\int_{\mathbb{R}} z^2\frac{3}{4}(1-z^2)dzI(|z|\le 1)\right)^{2/5} \\
&= \left(||K_1(z)||_2^2\right)^{4/5}\left(\mu_2(K_1(z))\right)^{2/5}
\end{aligned}$$

# MA 750: HW2

*Benjamin Draves*

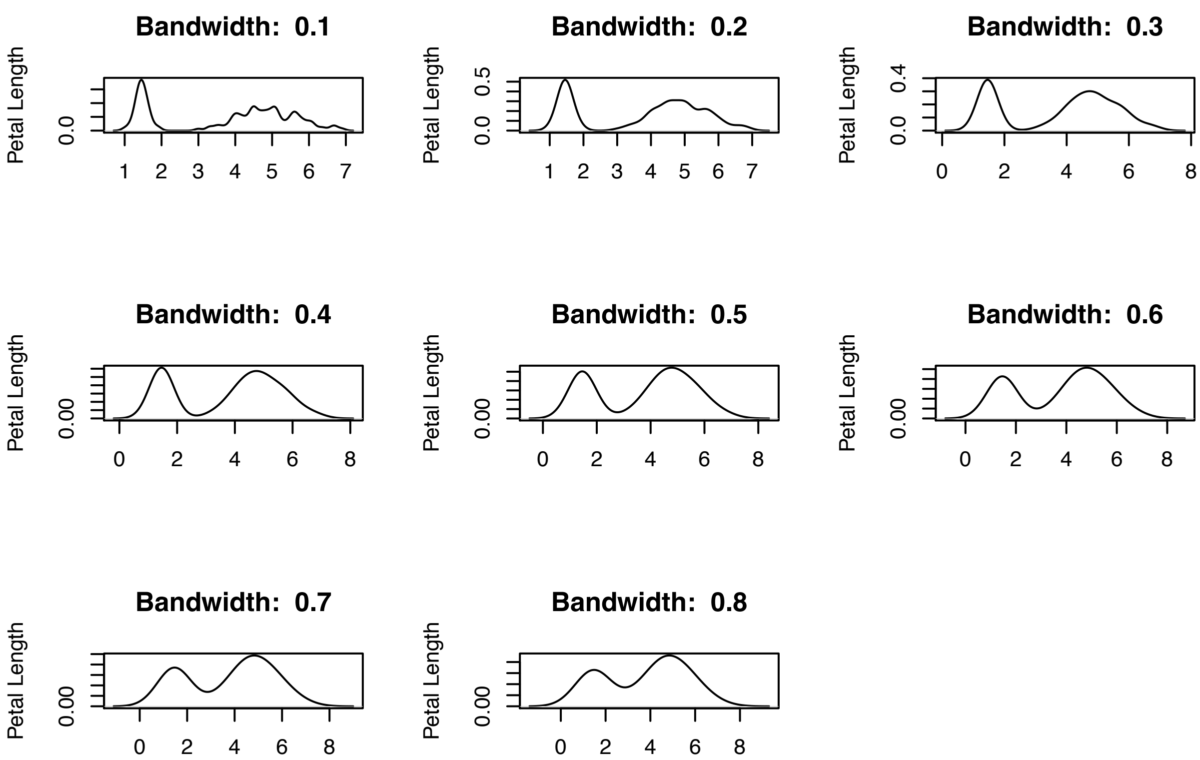
*October 5, 2017*

**Exercise 2.4**

```r
# take a peak at the data
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

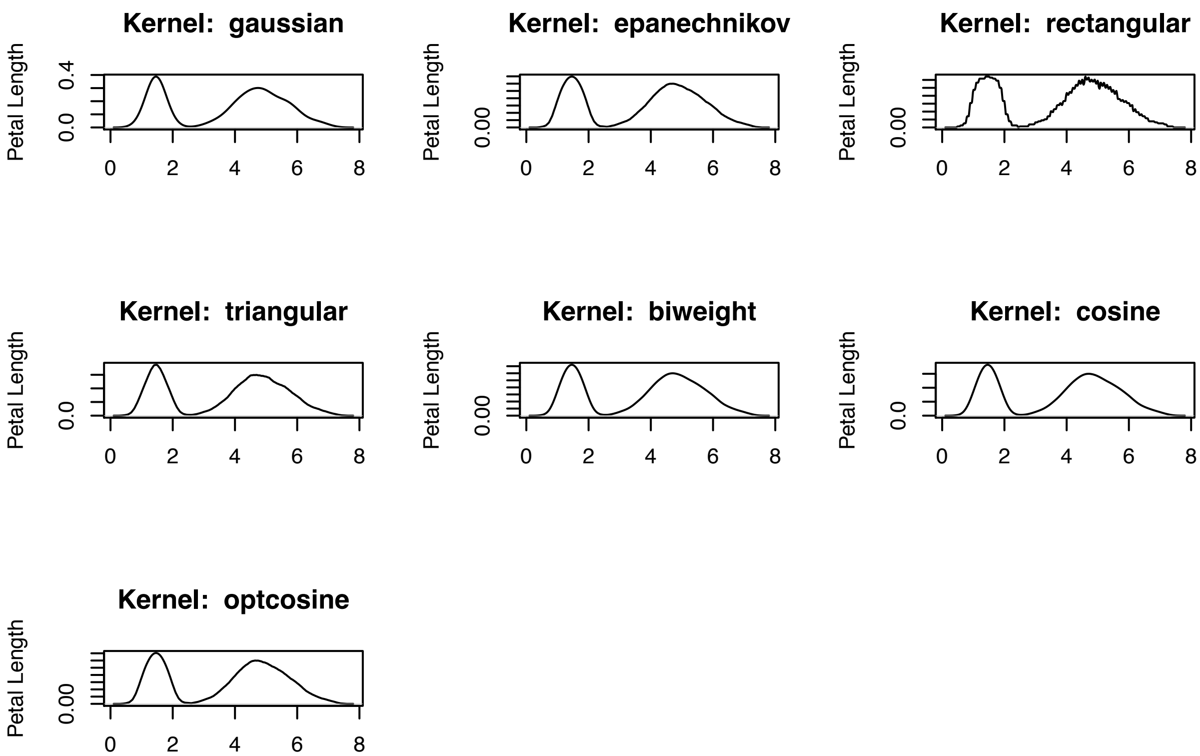
```r
# Take a look at multiple bandwidths
par(mfrow = c(3, 3))
for (h in seq(0.1, 0.8, 0.1)) {
    plot(density(iris$Petal.Length, bw = h, kernel = "gaussian"),
        xlab = "", ylab = "Petal Length", main = paste("Bandwidth: ",
            h))
}
```



It appears that bandwidths of $h > .4$ oversmooth the data. In fact, we see at $h \approx 0.45$ that the mode changes from the first feature to the second. This behaviors shows how bandwidth selection can dramatically change the results of this sommthing method. It appears that both $h = 0.1, 0.2$ undersmooth the data with several rigid edges. A bandwidth of $h = 0.3$ seems to best smooth the data - preserving local features while smoothing the data.

```r
# Take a look at multiple kerenels
kernels = c("gaussian", "epanechnikov", "rectangular", "triangular",
    "biweight", "cosine", "optcosine")

par(mfrow = c(3, 3))
for (kern in kernels) {
    plot(density(iris$Petal.Length, bw = 0.3, kernel = kern),
        xlab = "", ylab = "Petal Length", main = paste("Kernel: ",
            kern))
}
```




This chart shows how little effect the choice of kernel has on our model results. Aside from the rectangular kernel, all charts appear *relatively* similar. The triangular kernel may also fail to smooth the data appropriately while the optcosine and cosine may othersmooth the data. In my opinion, the Gaussian and Epanechnikov kernels preform the best. Although similar, I would choose the Epanechnikov kernel over all other kernels.