Stat 464, HW 10

1 In class

In this homework, we will focus on creating both histograms and kernel density estimators. Just for class, let's create some simulated data.

```
x=rchisq(50,5)
```

In `hw10functions.R` you will find the function `hist2(x,h)` which will plot a histogram as we've defined it in class. (`h` is the width of each bar.)

To find the kernel density estimator, you can use the command

```
density(x,bw=1,kernel="g")
```

to find the density estimator for `h` is one and using a Gaussian kernel. Another possibility is to use an Epanechnikov kernel

$$K(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{x^2}{5}\right)$$

for $-\sqrt{5} < x < \sqrt{5}$ by changing the argument to `kernel="e"` To actually plot this estimator, you use

```
plot(density(x,bw=1,kernel="g"))
```

You can have the density estimation minimize the estimated risk ($J(h)$ from class) by using

```
density(x,bw="bcv",kernel="g")
```

(Don't worry about any warnings.)

2 Homework

1. Input the data from exercise one of chapter 10; this is the average March temperatures for Kansas City from 1961 to 1990. (Use `x=source("temp.txt")`.) Look at histograms for different `h` values. By eye, find one histogram which has too much variance. Find one histogram which has too much bias. Find one histogram that you think strikes a good balance between the two.

2. Use the estimated risk to select an appropriate `h` for the histogram for the temperature data. Recall that the estimated risk for a histogram is given as

$$\hat{J}(h) = \frac{2}{(n-1)h} - \frac{n+1}{(n-1)h} \sum_{i=1}^{m} \hat{p}_j^2$$

   You can obtain the $\hat{p}_j$ using the function `histp(x,h)`

3. Use the temperature data and find the kernel density estimation (use a Gaussian kernel). By eye, find an example of too much variance. Find an example of too much bias. Find an example that you think strikes a good balance.

4. Find the density estimation where the bandwidth is automatically found using estimated risk. What do you think about minimizing the estimated risk? Do you think it is fairly accurate?