

1. a) I will use SURE to estimate the risk of the generalized James-Stein estimator. Let

$$\hat{\theta}^{JS,v} = Z - \frac{(Z - v)(n - 2)\sigma^2}{\sum_{i=1}^n (Z_i - v)^2} = Z + g(Z)$$

Accordingly,

$$\begin{aligned} D_i &= \frac{\partial g_i}{\partial z_i} \\ D_i &= -(n - 2)\sigma^2 \left( \frac{1}{\sum_{i=1}^n (z_i - v)^2} - \frac{(z_i - v)^2}{(\sum_{i=1}^n 2(z_i - v)^2)^2} \right) \\ \sum_{i=1}^n D_i &= -(n - 2)\sigma^2 \left( \frac{n}{\sum_{i=1}^n (z_i - v)^2} - \frac{2 \sum_{i=1}^n (z_i - v)^2}{(\sum_{i=1}^n (z_i - v)^2)^2} \right) \\ \sum_{i=1}^n D_i &= -(n - 2)\sigma^2 \left( \frac{n - 2}{\sum_{i=1}^n (z_i - v)^2} \right) \\ \sum_{i=1}^n D_i &= -\frac{(n - 2)^2 \sigma^2}{\sum_{i=1}^n (z_i - v)^2} \end{aligned}$$

Plugging this into the SURE formula, we get that

$$\begin{aligned} \hat{R}(Z) &= n\sigma^2 + 2\sigma^2 \sum_{i=1}^n D_i + \sum_{i=1}^n g_i^2 \\ \hat{R}(Z) &= n\sigma^2 - 2 \frac{(n - 2)^2 \sigma^4}{\sum_{i=1}^n (Z_i - v)^2} + \sum_{i=1}^n \left( \frac{-(Z_i - v)(n - 2)\sigma^2}{\sum_{i=1}^n (Z_i - v)^2} \right)^2 \\ \hat{R}(Z) &= n\sigma^2 - 2 \frac{(n - 2)^2 \sigma^4}{\sum_{i=1}^n (Z_i - v)^2} + \frac{(n - 2)^2 \sigma^4 \sum_{i=1}^n (Z_i - v)^2}{(\sum_{i=1}^n (Z_i - v)^2)^2} \\ \hat{R}(Z) &= n\sigma^2 - 2 \frac{(n - 2)^2 \sigma^4}{\sum_{i=1}^n (Z_i - v)^2} + \frac{(n - 2)^2 \sigma^4}{\sum_{i=1}^n (Z_i - v)^2} \\ \hat{R}(Z) &= n\sigma^2 - \frac{(n - 2)^2 \sigma^4}{\sum_{i=1}^n (Z_i - v)^2} \end{aligned}$$

Since the estimate is unbiased,

$$R(\hat{\theta}^{JS,v}, \theta) = E(\hat{R}(Z)) = n\sigma^2 - (n - 2)^2 \sigma^4 E \left( \frac{1}{\sum_{i=1}^n (Z_i - v)^2} \right)$$

Since  $Z_i - v \sim N(\theta_i - v, \sigma^2)$ , we can use a result from the book that

$$E \left( \frac{1}{\sum_{i=1}^n (Z_i - v)^2} \right) \geq \frac{1}{(n - 2)\sigma^2 + \|\theta - v\|^2}$$

This gives us that

$$\begin{aligned}
R(\hat{\theta}^{JS,v}, \theta) &\leq n\sigma^2 - \frac{(n-2)^2\sigma^4}{(n-2)\sigma^2 + \|\theta - v\|^2} \\
R(\hat{\theta}^{JS,v}, \theta) &\leq \frac{n\sigma^2((n-2)\sigma^2 + \|\theta - v\|^2) - (n-2)^2\sigma^4}{(n-2)\sigma^2 + \|\theta - v\|^2} \\
R(\hat{\theta}^{JS,v}, \theta) &\leq \frac{n(n-2)\sigma^4 + n\sigma^2\|\theta - v\|^2 - (n-2)^2\sigma^4}{(n-2)\sigma^2 + \|\theta - v\|^2} \\
R(\hat{\theta}^{JS,v}, \theta) &\leq \frac{2\sigma^2(n-2)\sigma^2 + (2+n-2)\sigma^2\|\theta - v\|^2}{(n-2)\sigma^2 + \|\theta - v\|^2} \\
R(\hat{\theta}^{JS,v}, \theta) &\leq 2\sigma^2 + \frac{(n-2)\sigma^2\|\theta - v\|^2}{(n-2)\sigma^2 + \|\theta - v\|^2}
\end{aligned}$$

Because

$$1 < \frac{\|\theta - v\|^2}{(n-2)\sigma^2 + \|\theta - v\|^2}$$

We get that, for  $n \geq 3$ ,

$$R(\hat{\theta}^{JS,v}, \theta) \leq 2\sigma^2 + \frac{(n-2)\sigma^2\|\theta - v\|^2}{(n-2)\sigma^2 + \|\theta - v\|^2} < 2\sigma^2 + (n-2)\sigma^2 = n\sigma^2$$

$R(\hat{\theta}^{JS,v}, \theta)$  will be smallest when  $\frac{\|\theta - v\|^2}{(n-2)\sigma^2 + \|\theta - v\|^2}$  is smallest, which in turn is when  $\|\theta - v\|^2$  is smallest. Thus,  $v = \bar{\theta}$ , the mean of  $\theta$ , will result in the smallest risk.

b) Let  $e \sim N(0, \sigma^2)$ . For  $n = 1$ , we can calculate the risk of  $\hat{\theta}^{a,b}$ :

$$\begin{aligned}
R(\hat{\theta}^{a,b}, \theta) &= E((aZ + b - \theta)^2) \\
&= E((ae + a\theta - \theta + b)^2) \\
&= E(a^2e^2 + a^2\theta^2 + \theta^2 + b^2 + 2a^2e\theta - 2ae\theta + 2aeb - 2a\theta^2 + 2ba\theta - 2b\theta) \\
&= a^2\sigma^2 + a^2\theta^2 + \theta^2 + b^2 - 2a\theta^2 + 2ba\theta - 2b\theta \\
&= a^2\sigma^2 + (a-1)^2\theta^2 + b^2 + 2(a-1)b\theta \\
&= a^2\sigma^2 + ((a-1)\theta + b)^2
\end{aligned}$$

(i)  $a = 0$  is admissible, since this is the best predictor when  $b = \theta$ , as we can see from the risk:

$$R(\hat{\theta}^{0,b}, \theta) = \theta^2 + b^2 + 2(-1)b\theta = 2\theta^2 - 2\theta^2 = 0$$

If a estimator  $\tilde{\theta}$  is not uniformly  $\theta = b$ , then it will have positive risk, so  $R(\hat{\theta}^{0,b}, \theta) < R(\tilde{\theta}, \theta)$ , meaning  $\hat{\theta}^{0,b}$  must be admissible.

(ii)  $\hat{\theta}^{a,b}$  for  $a < 0$  is not admissible, by this line of reasoning. If  $a < 0$ , the  $0 < \left(\frac{1}{a-1}\right)^2 < 1$ , so

$$\begin{aligned}
R(\hat{\theta}^{a,b}, \theta) &= a^2\sigma^2 + ((a-1)\theta + b)^2 \\
R(\hat{\theta}^{a,b}, \theta) &> \left(\frac{1}{a-1}\right)^2 \left(a^2\sigma^2 + ((a-1)\theta + b)^2\right) \\
R(\hat{\theta}^{a,b}, \theta) &> \left(\frac{a}{a-1}\right)^2 \sigma^2 + \left(\theta + \frac{b}{a-1}\right)^2 \\
R(\hat{\theta}^{a,b}, \theta) &> \left(-\theta - \frac{b}{a-1}\right)^2 \\
R(\hat{\theta}^{a,b}, \theta) &> R(\hat{\theta}^{0, \frac{-b}{a-1}}, \theta)
\end{aligned}$$

(iii)  $\hat{\theta}^{a,b}$  for  $a > 1$  is not admissible since, for all  $\theta$ ,

$$R(\hat{\theta}^{a,b}, \theta) = a^2 \sigma^2 + ((a-1)\theta + b)^2 > \sigma^2 + b^2 = R(\hat{\theta}^{1,b}, \theta)$$

(iv) + If  $a = 1$  and  $b \neq 0$ , then  $\hat{\theta}^{a,b}$  is not admissible, since, for all  $\theta$

$$R(\hat{\theta}^{1,b}, \theta) = \sigma^2 + b^2 > \sigma^2 = R(Z, \theta)$$

When  $n \geq 3$ , part a) proves that, for all  $\theta$ ,

$$R(Z, \theta) = n\sigma^2 > R(\hat{\theta}^{JS,v}, \theta)$$

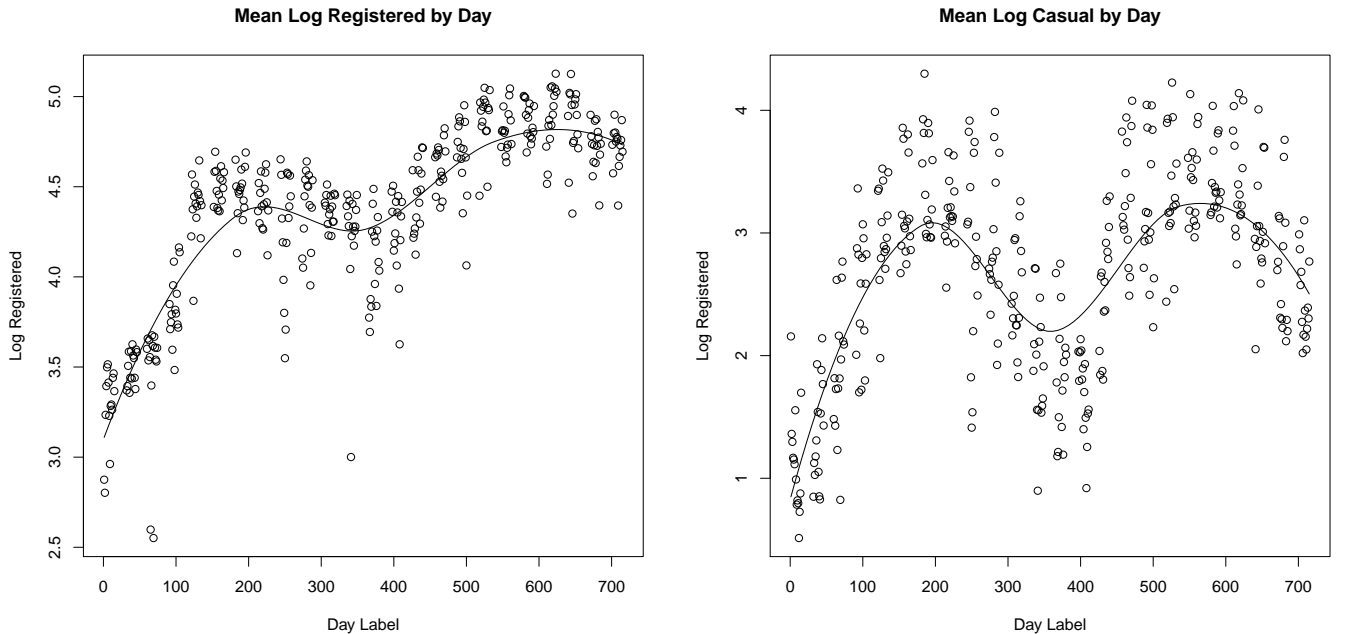
So  $Z = \hat{\theta}$  is not admissible. As (i) showed, for any particular  $\theta$ , A linear estimate like  $\hat{\theta}^{a,b}$  is not admissible if  $a < 0$ . Since the generalized James-Stein estimator is of this form, and can have a negative  $a$ , we can make a strictly improved version of it. This estimator would be equal to the generalize James-Stein estimator when it has a non-negative  $b$ , and  $\hat{\theta}^{0, \frac{-b}{a-1}}$ , for the  $b$  and  $a$  from the generalized James-Stein estimator when it wasn't. As proved in (ii), this estimator would have the same risk when  $a \geq 0$ , and would be strictly less risk when it wasn't. Thus, we can conclude the generalized James-Stein estimator is not admissible.

c) Running the three estimates (all results in radians), I got a total risk of 17.867 with the MLE, 16.636 with the normal James-Stein estimator ( $v = 0$ ), and 5.280 with the generalized James-Stein estimator with  $v = \bar{Z} = -3.317$ . Unsurprisingly with a data set that is not centered around 0, we see a huge improvement with the generalized James-Stein method set to the population mean over the MLE and original James-Stein estimators.

d)  $\frac{9}{45}$  is almost certainly a less risky estimate. This is because there is no reason to expect that the portion of imported cars seen in Chicago and 18 players batting averages have the same variance. Since this violates an assumption of the James-Stein estimator, it would be inappropriate to use it.

2. a) For both log counts, I used a linear model that included the season as factor, working day as a factor, the interaction of season and working day as a factor, weather as factor, day label as a continuous variable, and hours separation from 5pm as a continuous variable. In total, I got a RMSLE of 1872.738 for the registered users and 2163.457 for the casual non-registered users.

b) These are the fits of the two local linear regressions on the mean hourly transformed counts by day:



With the new local linear regression and the linear model predicted on its residuals, I got RMSLE of 1828.168 for the registered users and 1875.078 for the casual non-registered users.

- c) I ran a generalized additive model with just smoothed functions for daylabel, humidity, temp, windspeed, and hour. The end result was that I got a RMSLE of 1352.043 for registered customers and 1326.217 for casual customers.
- d) For the solution data set I turned in, I predicted the total counts with an additive linear model predicting the log of the total counts with smoothed daylabel, humidity, temp, wind speed, hour, and categorical of working day, and weather interacted with season.

3. Starting with the bias:

$$\begin{aligned}
bias(\hat{p}_n(x_0, y_0)) &= p(x_0, y_0) - E(\hat{p}_n(x_0, y_0)) \\
&= p(x_0, y_0) - \frac{1}{h^2} \iint K\left(\frac{x_0 - t}{h}\right) K\left(\frac{y_0 - q}{h}\right) p(q, t) dt dq \\
&= \frac{1}{h^2} \iint K\left(\frac{x_0 - t}{h}\right) K\left(\frac{y_0 - q}{h}\right) (p(q, t) - p(x_0, y_0)) dt dq \\
&\leq \frac{1}{h^2} \iint K\left(\frac{x_0 - t}{h}\right) K\left(\frac{y_0 - q}{h}\right) L(|t - x_0|^\beta + |q - y_0|^\beta) dt dq \\
&\leq \iint K(t) K(q) L(|ht|^\beta + |hq|^\beta) dt dq \\
&\leq 4Lh^\beta \iint_{\mathbb{R}^+ \times \mathbb{R}^+} K(t) K(q) (t^\beta + q^\beta) dt dq \\
&\leq h^\beta c_1
\end{aligned}$$

Since each  $X_i$  is iid and the kernel estimator is linear, we can decompose the variance:

$$\begin{aligned}
\text{Var}(\hat{p}_n(x_0, y_0)) &= \frac{1}{nh^4} \text{Var}\left(K\left(\frac{X_i - x_0}{h}\right) K\left(\frac{Y_i - y_0}{h}\right)\right) \\
&= \frac{1}{nh^4} E\left(\left(\iint K\left(\frac{X - x_0}{h}\right) K\left(\frac{Y - y_0}{h}\right) p(t, q) - K\left(\frac{x_0 - t}{h}\right) K\left(\frac{y_0 - q}{h}\right) p(t, q) dt dq\right)^2\right) \\
&\leq \frac{1}{nh^4} E\left(\left(\iint K\left(\frac{X - t}{h}\right) K\left(\frac{Y - q}{h}\right) p(t, q) dt dq\right)^2\right)
\end{aligned}$$

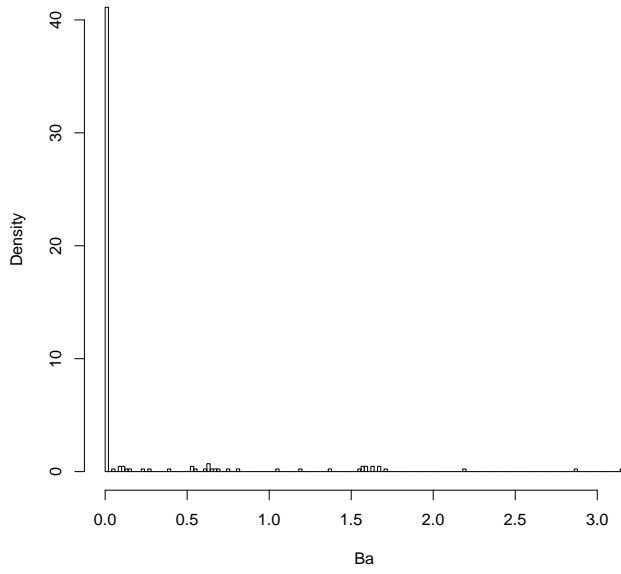
And here is where I am stuck. However, if we had an expression for variance, then the mean squared error would be

$$bias(\hat{p}_n(x_0, y_0))^2 + \text{Var}(\hat{p}_n(x_0, y_0)) \leq h^{2\beta} c_1 + \text{Var}(\hat{p}_n(x_0, y_0))$$

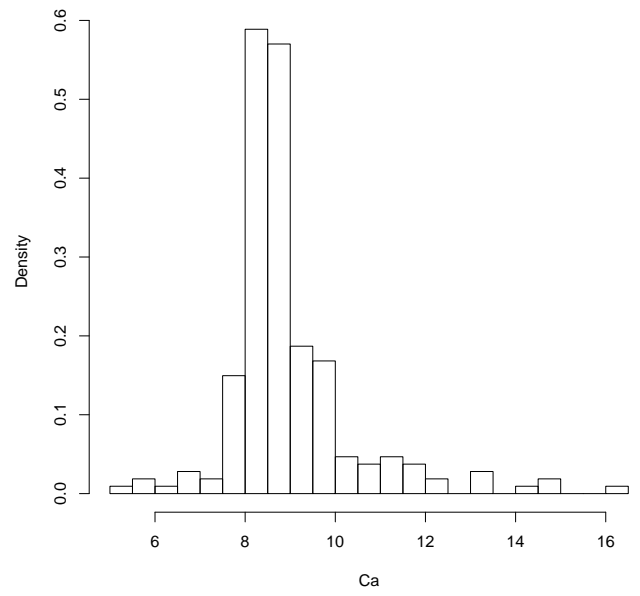
Differentiating the right side of that with respect to  $h$  and setting it to 0, we would find the optimal  $h$ . That would have a leading term of the order  $h^{2\beta-1}$ , at least assuming the bias has the larger  $h$  power. Then, we would expect to see  $h$  shrink with  $cn^{\frac{-1}{2\beta+1}}$ , dictating  $R$  converge at a rate of  $O(n^{\frac{2\beta}{2\beta+1}})$ .

- 4. a) Here are the histograms of the variable densities. The number of bins was chosen between 1 and 200 based on cross validation score.

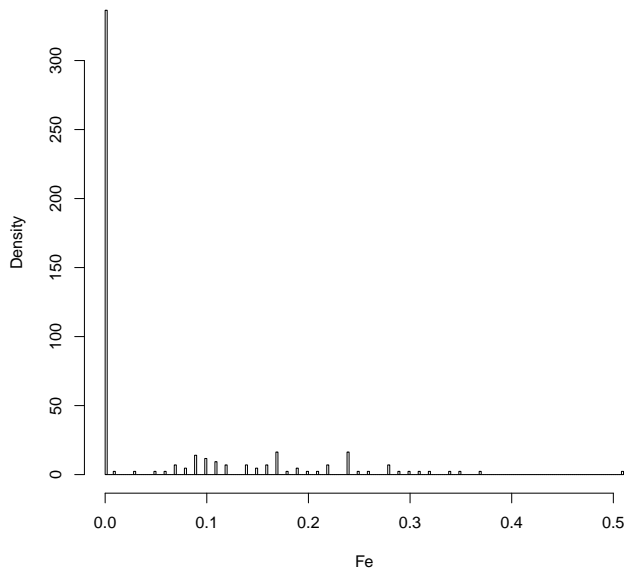
**Histogram of Ba, 200 bins**



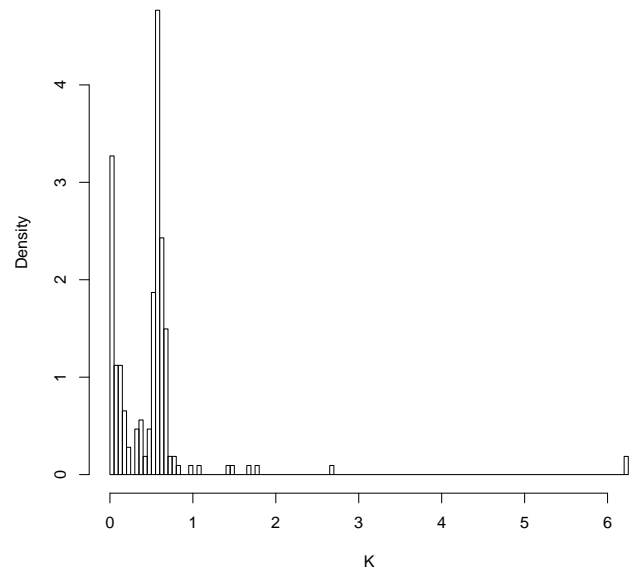
**Histogram of Ca, 22 bins**

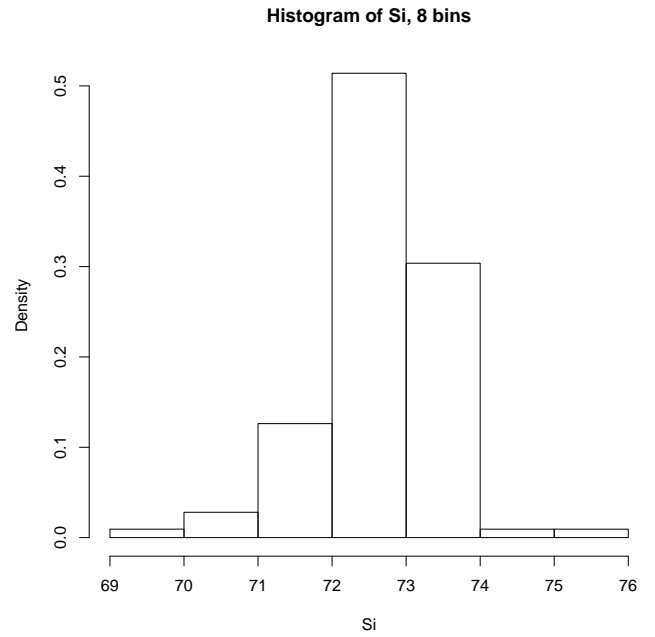
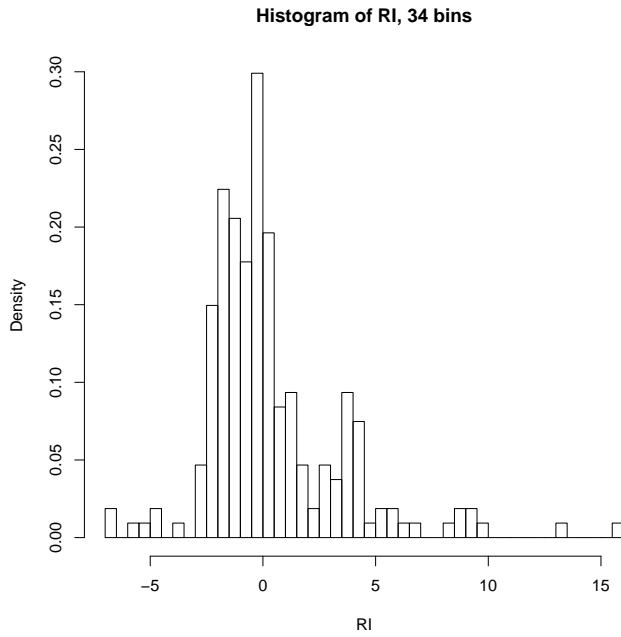
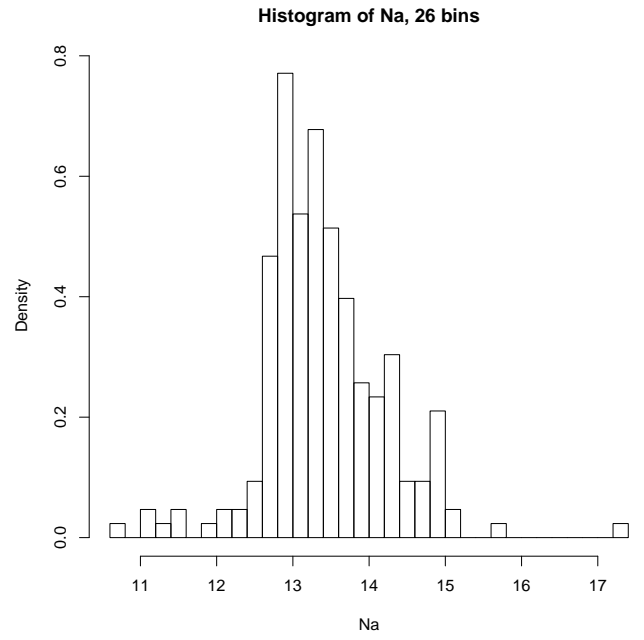
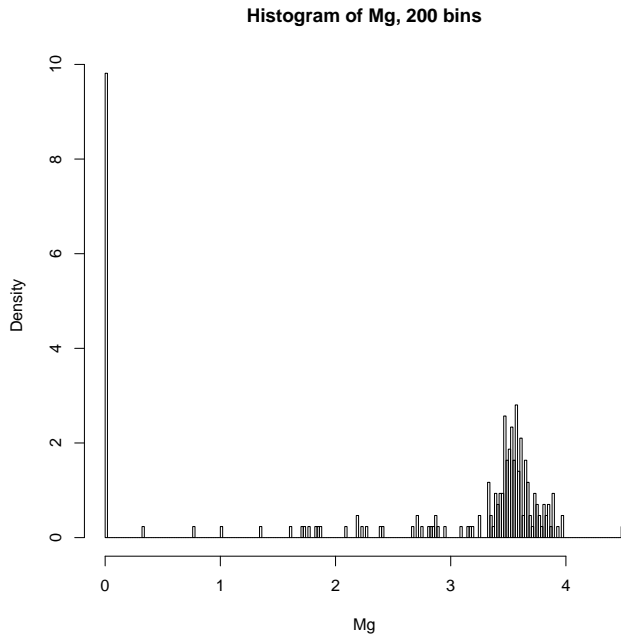


**Histogram of Fe, 200 bins**



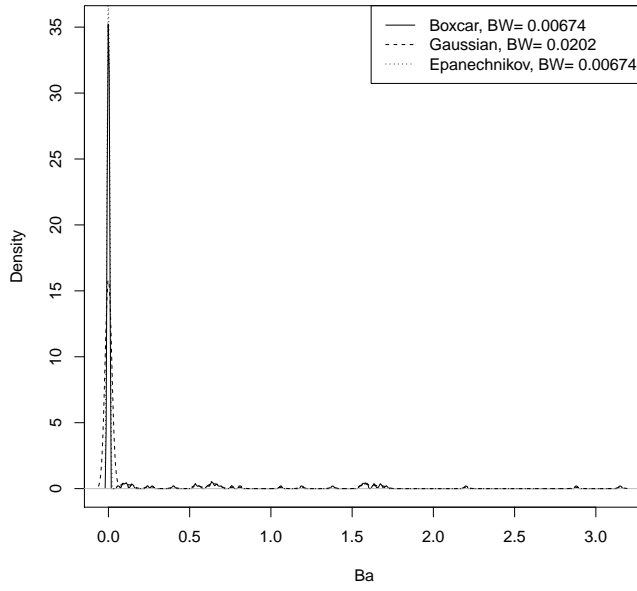
**Histogram of K, 200 bins**



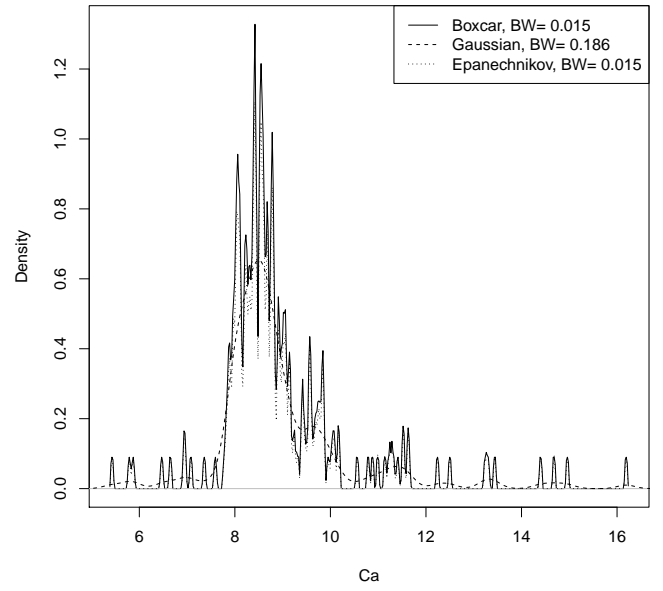


- b) These are the density estimates with bandwidth picked by cross validation. The Gaussian Kernel's bandwidth was picked with R's built in unbiased cross validation tool, while the other two I calculated manually on a series of points around the optimal Gaussian bandwidth, picking the smallest estimated risk. In general, the Gaussian seemed to go with the higher bandwidth, producing the smoother estimate, but that may be in part due to a failure in my implementation of cross validation. Certainly it is smoother than the other two kernels by construction, since they are 0 except over a closed interval.

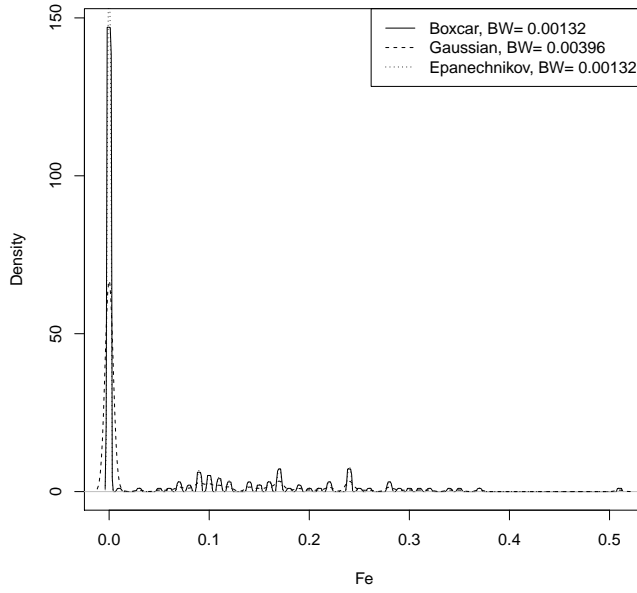
**Ba Density Estimates**



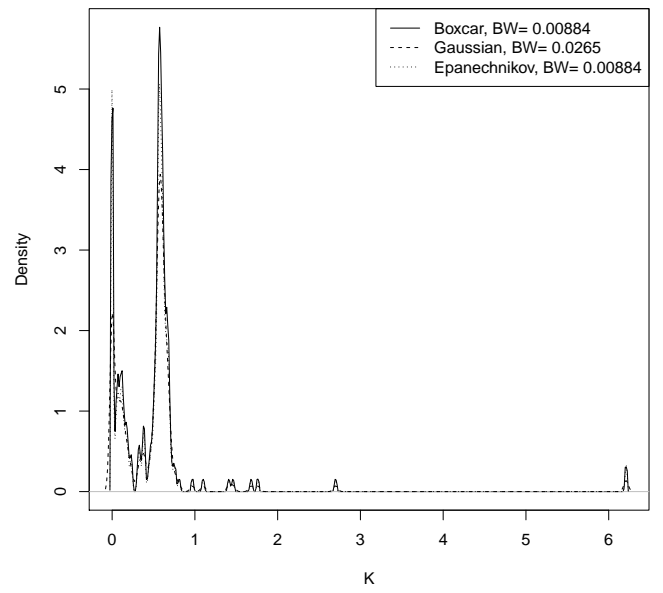
**Ca Density Estimates**

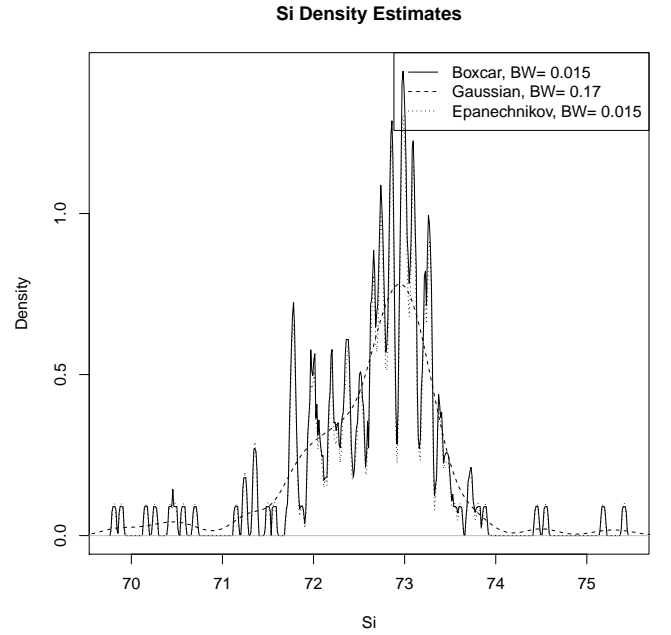
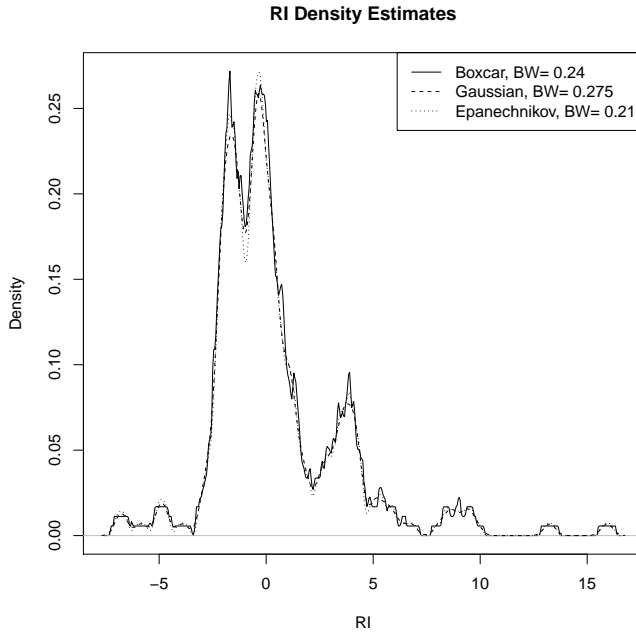
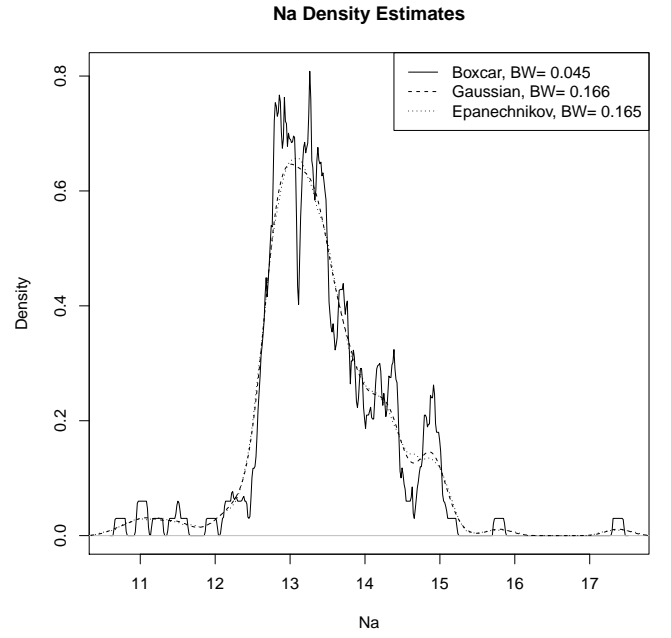
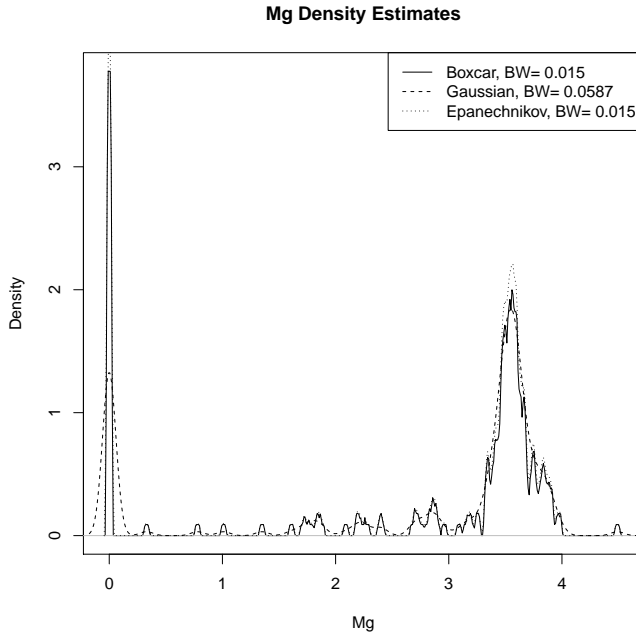


**Fe Density Estimates**



**K Density Estimates**





- c) Cross validation seems to do poorly with variables with one or a few large point masses. We see this several times with a large number of points being at 0. This is because cross validation rewards the estimator for having an extremely sharp peak around these point masses, even though the rest of the distribution suffers. One might view these as almost two separate distributions. A better way to estimate these sort of variables would be to predict the point mass separately, remove those points, and then run the density estimate on the rest of the distribution.