Solutions for Homework 3 Stat 550

Dr. Scott

Chapter 3 40 Points

Instructions: We will discuss generally in class before due date. You can work in groups, but turn in your own solutions.

3.4: To compare to the Normal rule

$$\begin{aligned} h^* &= \left[\frac{6}{n\,R(f')} \right]^{1/3} \\ &= 3.4908\,\sigma\,n^{-1/3} \qquad \text{for } N(\mu,\sigma^2) \text{ data} \end{aligned}$$

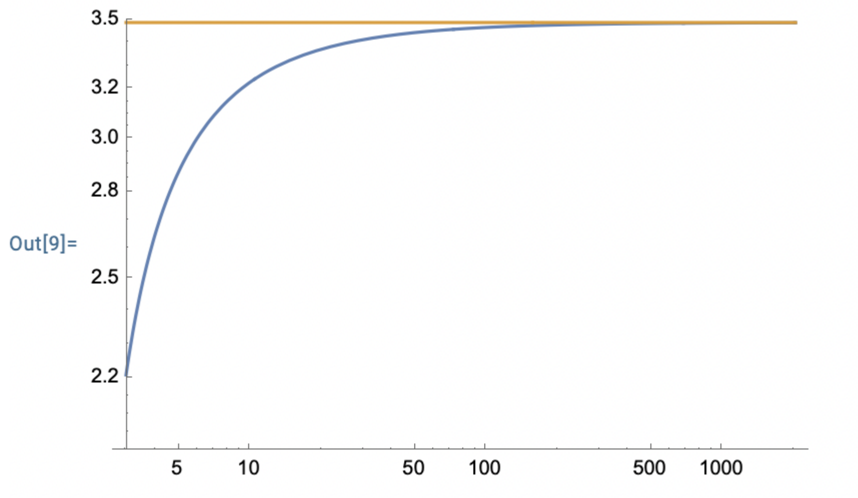
we will want to insert the standard deviation of the pdf in the appropriate place. Everything else will be compared to the constant 3.4908. For example, if we focus on the T_p density, then

$$\sigma_p = \sqrt{\frac{p}{p-2}}\,.$$

Use Mathematica to compute $c = R(f')$, then write as

$$h^* = \left[\frac{6}{c}\right]^{1/3}\;\frac{1}{\sigma_p}\,\sigma_p\,n^{-1/3}\,.$$

The text file `tp.ans.txt` contains the Mathematica code to compute the constant to be compared to 3.4908. The Gamma function will overflow if $p > 150$ or so. I discovered the PowerExpand function on the LogGamma, which expands the plotable range of p into the thousands; see the figure so generated below. Since the Cauchy is a T_1 pdf, this includes that case. Other pdf's similar.



3.5: Using

$$AMISE(h) = \frac{1}{nh} + \frac{1}{12} h^2 R(f')$$
$$h^* = \left[\frac{6}{n\,R(f')} \right]^{1/3}$$

we obtain

$$AMISE(h) = 6^{-1/3}\ [1+2]\ R(f')^{1/3}\, n^{-1/3}\,,$$

which clearly shows the 2:1 ratio of AIBS to AIV.

3.10: We use the exact MISE formula given after Eq (3.25),

$$MISE = \frac{1}{nh} - \frac{n+1}{nh} \sum_k p_k^2 + R(f)\,.$$

The true pdf is $f \sim U(0,1)$, so $R(f) = 1$.

(a.) With $t_k = kh$ and m bins, $h = 1/m$ and $p_k = 1/m$ as well. Then

$$\sum_{k=0}^{m-1} p_k^2 = m \times \left(\frac{1}{m} \right)^2 = \frac{1}{m}\,,$$

which leads to $(m-1)/n$.

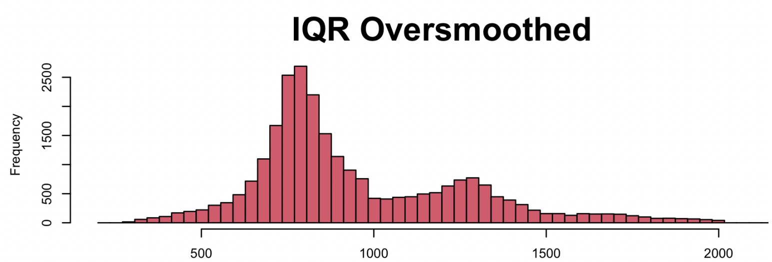
(b.) With the misaligned bins, the first and last bins have probability of $1/(2m)$ while the remaining bins have probability $1/m$ as above. Hence,

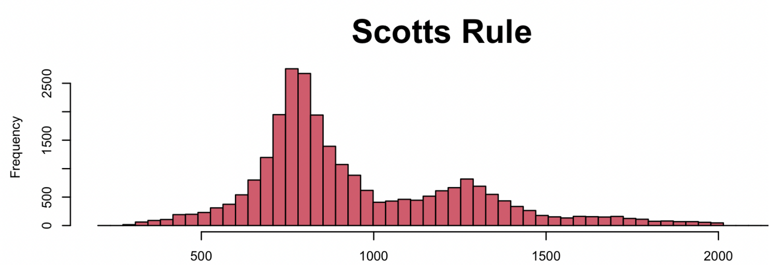
$$\sum_{k=0}^{m} p_k^2 = 2 \times \left(\frac{1}{2m}\right)^2 + (m-1) \times \left(\frac{1}{m}\right)^2 = \frac{1}{2m^2} + \frac{m-1}{m^2}\,,$$

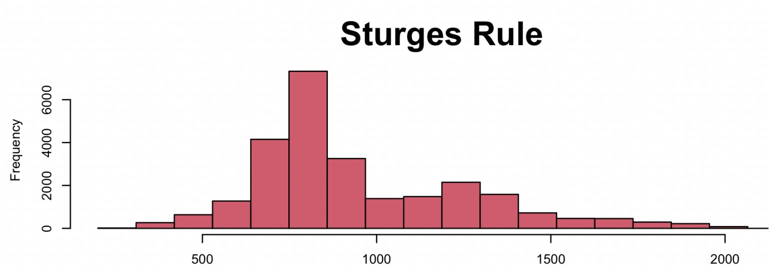
etc.

3.14: Create a vector x by concatenating uniform samples for all 172 bins:

```
xb = scan("lrl.dat",n=172); x = NULL; h=10
for(i in 1:172) { ti=280+i*h; x=c(x,runif(xb[i],ti, ti+10 ))}
iqr = IQR(x)      # 404.15
hOS = 2.603 * iqr * length(x)**(-1/3)     # 35.62
hist(x, seq(200,2200,hOS), col=2, main="IQR Oversmoothed",xlab="")
```







3.15: (Bonus) See the notes for the Lagrange derivation. Our initial solution neglected to insist that the polynomial not only match values at the 4 knots, but also the first derivatives should match. The 2 Mathematica files `iqr-os.txt` and `iqr-os-fix.txt` show both solutions,