Question 1

In Questions 1.2 up to and including 1.5, we reset the seed to 123 each time.

Question 1.1

In the data set x the value 0.47 appears twice. Hence, the probability of a random bootstrap sample of 10 data points containing 0.47 10 times is $(\frac{1}{5})^{10}$. Any of the eight other values other than 0.47 appears only once in x, such that the probability of a random bootstrap sample of 10 data points containing such value 10 times is $(\frac{1}{10})^{10}$. To conclude, the probability that a random bootstrapped sample from x consists of 10 repeated values is given by:

$$p = 8\left(\frac{1}{10}\right)^{10} + \left(\frac{1}{5}\right)^{10}.$$

Question 1.2

Figure 1 shows five bootstrapped samples of x.

```
     [,1]  [,2] [,3]  [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 0.23  0.23 2.50  0.20 0.59 0.47 0.47 0.59 2.03  2.50
[2,] 0.47  0.23 2.03  2.03 2.03 0.23 1.01 2.50 0.66  2.50
[3,] 2.03  0.23 0.47 -0.76 0.66 0.47 2.50 0.66 2.03  2.03
[4,] 2.50  0.66 0.47  0.66 0.47 0.59 2.03 0.20 0.47  1.01
[5,] 0.20 -0.76 2.03  2.03 0.59 0.47 2.03 2.50 0.47  0.59
```

Figure 1: 5 bootstrapped samples from x

Question 1.3

Given our parameter of interest $\theta(F) = \mathbb{E}_F[X]/\mathbb{E}_F[Y]$, where expectations are taken over the actual distribution function F, its plug-in estimator is given by $\hat{\theta} = \theta(\hat{F})$, where expectations are taken over the empirical distribution function $\hat{F}$. Hence, we obtain the following derivation which shows that $\hat{\theta}$ equals the division of the mean of x by the mean of y:

$$\hat{\theta} = \frac{\mathbb{E}_{\hat{F}}[X]}{\mathbb{E}_{\hat{F}}[Y]} = \frac{\sum_{i=1}^{n} x_i P[X = x_i]}{\sum_{i=1}^{n} y_i P[Y = y_i]} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} y_i} = \frac{\bar{x}}{\bar{y}}$$

Question 1.4

Having bootstrapped 200 samples of X and Y, we have computed the estimate $\hat{\theta}^* = \bar{x}/\bar{y}$ for θ for each of the 200 samples. Figure 2 shows the histogram of these 200 values bootstrapped estimates.

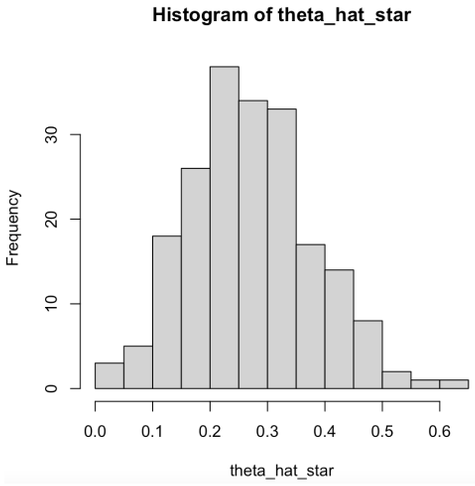
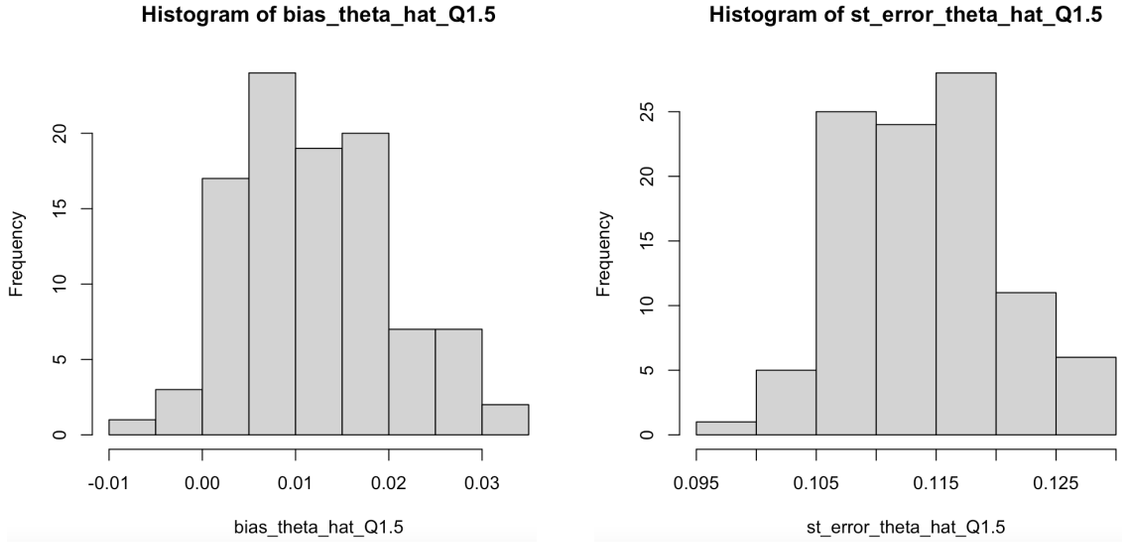


Figure 2: Histogram of $\hat{\theta}^*$s

We can compute the bias and standard error of $\hat{\theta}$ as follows. The bias of $\hat{\theta}$ is obtained by subtracting $\hat{\theta}$, the initial estimate of θ based on the given data set, from the sample mean of $\hat{\theta}^*$s, given in Figure 2. This yields a bias of 0.0229. The standard error estimate is equal to the sample standard deviation of the $\hat{\theta}^*$s, which is 0.1093.

Question 1.5

We repeat the procedure of Question 1.4 100 times in order to obtain 100 estimates of the bias and of the standard error of $\hat{\theta}$ based on the 100 different samples of the $\hat{\theta}^*$s. Figure 3 shows a histogram of the estimated biases and standard errors of $\hat{\theta}^*$.



(a) Bias of $\hat{\theta}^*$

(b) Standard error $\hat{\theta}^*$

Figure 3: Bias and standard error of $\hat{\theta}^*$

We would say that $\hat{\theta}^*$ does not look like an unbiased estimator. By delving a bit further into the results, the mean of the biases equals 0.0122, whereas the average standard error equals 0.114. Since the average standard error is around 10 times as large in magnitude as the average bias, we could however conclude that the $\hat{\theta}$ seems unbiased.

Question 1.6

There are two reasons for the relatively large bias in the estimation.

The first reason is that the empirical distribution function $\hat{F}$ generally does not approximate the true distribution F well. As we only have a sample size of 10 observations, there is no guarantee that the empirical distribution function approximates the true distribution.

The second reason is the fact that this estimator is not robust. The plug-in estimator $\hat{\theta} = \frac{x}{y}$ does not work well in case of outliers or data with heavy tails. In fact, the t-distribution can be characterized by its heavy tails, causing the estimator to be highly unstable in this case.

Question 2

In Questions 2.2 up to and including 2.5, we reset the seed to 123 each time.

Question 2.1

We have modelled the data using the linear regression $y = \beta_0 + \beta_1 x + \epsilon$. The OLS estimates for β_0 and β_1 are -0.759 and 5.28, respectively. Moreover, the standard errors obtained using the formula for least-squares regression are 1.091 and 0.109, respectively.

Question 2.2

We have bootstrapped a set of $(x_1^*, y_1^*), ..., (x_{500}^*, y_{500}^*)$ using parametric bootstrap. The procedure is depicted in the following algorithm:

Algorithm 1 Parametric bootstrap

1: Take random sample of 500 with replacement of residuals from Question 2.1.
2: $x^* = \mathbf{x}$
3: **for** $i = 1, 2, \ldots, 500$ **do**
4: Calculate $y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \epsilon_i^*$
5: **end for**

First bootstrap a random sample of 500 with replacement from the residuals obtained in Question 2.1, and denote these by ϵ_i^*. Hereafter, set x^* equal to initial $\mathbf{x}$. Finally, iterate over i with $i = 1, ..., 500$ and calculate for each i:

$$y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \epsilon_i^*.$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS bootstrapped estimates computed in Question 2.1. Then, this set of bootstrapped values of x_i^* and y_i^* is used to obtain OLS bootstrapped estimates $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

Question 2.3

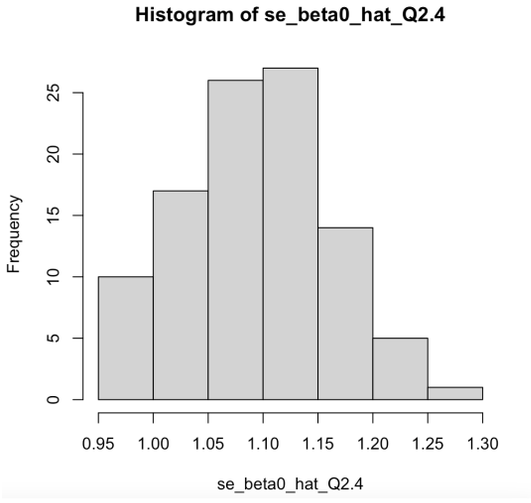
Using the parametric bootstrap procedure we have calculated the fitted coefficients β_0^* and β_1^*. These have values equal to -1.53 and 5.35, respectively.

Question 2.4

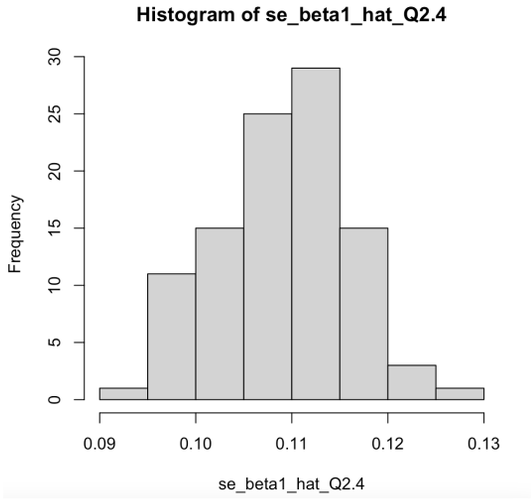
We have bootstrapped 100 sets of $(x_1^*, y_1^*), ..., (x_{500}^*, y_{500}^*)$ using parametric bootstrap resulting in 100 sets of estimates of β_0^* and β_1^*. The standard error of $\hat{\beta}$ is equal to 1.109 for $\hat{\beta}_0$ and 0.109 for $\hat{\beta}_1$, respectively.

Question 2.5

We have repeated Question 2.4 100 times to obtain 100 values of $\hat{se}^*(\hat{\beta})$. Figure 4 shows the histograms of standard errors of $\hat{\beta}^*$.



(a) Histogram of standard errors of $\hat{\beta}_0^*$



(b) Histogram standard errors $\hat{\beta}_1^*$

Figure 4: Histograms of standard errors of $\hat{\beta}^*$

Question 2.6

We choose to delve deeper into the data in order to propose an improvement of our bootstrap method. We inspect the dynamics of the data and observe correlation first- and third-order correlation between the data points of the dependent variable as displayed by the partial autocorrelations in Figure 5.

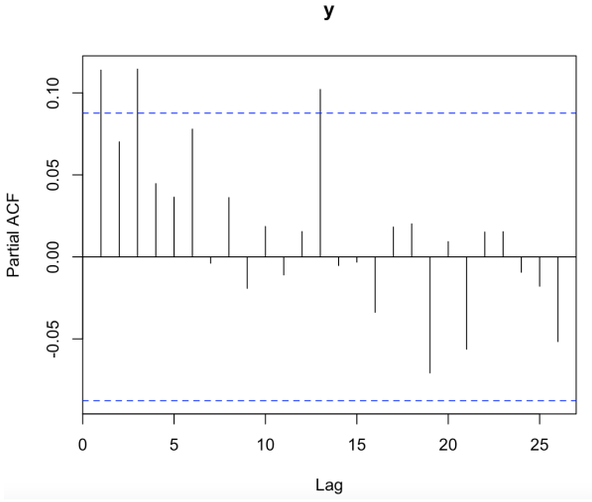


Figure 5: Partial autocorrelations in dependent variable y.

This suggests an AR(1) or AR(3) specification, in which 1 or 3 lags of the dependent variable are added as independent variables. For such time series specifications the straightforward method to bootstrap this is moving blocks bootstrap on the dependent variable. Choosing for an AR(3) specification, this method divides the set of dependent variables into blocks of 3. Then, we sample from these blocks with replacement yielding to a bootstrapped time series in which lags are introduced to the model.

Question 3

3.1

We have calculated $\mu_2(K)$ and $\|\|K\|\|_2^2$. The steps made for these calculations are shown below:

$$\begin{aligned}
\mu_2(K) &= \int s^2 K(s) ds \\
&= \int s^2 (\frac{3}{4})(1-s^2) ds \\
&= (\frac{3}{4})^2 (1-s^2) ds \\
&= (\frac{3}{4}) \int (s^2 - s^4) ds \\
&= (\frac{3}{4})^2 \left[\frac{1}{3} s^3 - \frac{1}{5} s^5 \right]_{-1}^{1} \\
&= \frac{3}{4}((\frac{1}{3} - \frac{1}{5}) - (-\frac{1}{3} + \frac{1}{5}) \\
&= \frac{1}{5}
\end{aligned}$$

$$\begin{aligned}
\|\|K\|\|_2^2 &= \int K^2(s) ds \\
&= \int (\frac{3}{4})^2 (1-s^2)^2 ds \\
&= (\frac{3}{4})^2 \int (1-s^2)^2 ds \\
&= (\frac{3}{4})^2 \int (1^2 + s^4 - 2s^2) ds \\
&= (\frac{3}{4})^2 \left[s + \frac{1}{5} s^5 - \frac{2}{3} s^3 \right]_{-1}^{1} \\
&= \frac{9}{16}((\frac{15}{15} + \frac{3}{15} - \frac{10}{15}) - (-\frac{15}{15} - \frac{5}{15} + \frac{10}{15})) \\
&= \frac{9}{16}((\frac{8}{15}) - (-\frac{8}{15})) \\
&= \frac{3}{5}
\end{aligned}$$

3.2

In addition to question 3.1, we have calculated the $AMISE\{\hat{f}_h\}$, as described:

$$AMISE\{\hat{f}_h\} = \frac{h^4}{4} \|\|f''\|\|_2^2 (\mu_2(K))^2 + \frac{1}{nh} \|\|K\|\|_2^2,$$

$$\|\|K\|\|_2^2 = \frac{3}{5},$$

$$\mu_2(K) = \frac{1}{5},$$

$$\|\|f''\|\|_2^2 = \sigma^{-5} \int \{\varphi''(u)\}^2 du = \sigma^{-5} \frac{3}{8\sqrt{\pi}}$$

$$AMISE\{\hat{f}_h\} = \frac{h^4}{4}(\sigma^{-5}\frac{3}{8\sqrt{\pi}})(\frac{1}{5})^2 + \frac{1}{nh}\frac{3}{5}$$
$$= \frac{3}{800\sqrt{\pi}}h^4\sigma^{-5} + \frac{1}{nh}\frac{3}{5}$$

$$\text{Bias square term } = \frac{3}{800\sqrt{\pi}}h^4\sigma^{-5}$$
$$\text{Variance term } = \frac{1}{nh}\frac{3}{5}$$

3.3

In Figure 6, we have plotted the Bias square term (red line), variance term (blue line) and the total AMISE (black line). From the figure, we can see that the optimal h, hence the point where the black curve is lowest is around 2.

To confirm this number, we take the derivative of the total AMISE and equal it to zero:

$$\frac{d}{dh}\left(\frac{3}{800\sqrt{\pi}}h^4\sigma^{-5} + \frac{1}{nh}\frac{3}{5}\right) = \frac{4*3}{32*800*\sqrt{\pi}}h^3 + \frac{-1}{nh^{-2}*0.6} = 0$$

$$\frac{3}{8*800*\sqrt{\pi}} * h^3 = \frac{0.6}{nh^2}$$

$$h^5 = \frac{8*800\sqrt{\pi}*6}{3000}$$

Solving for h yields $h_{opt} \approx 1.867$.

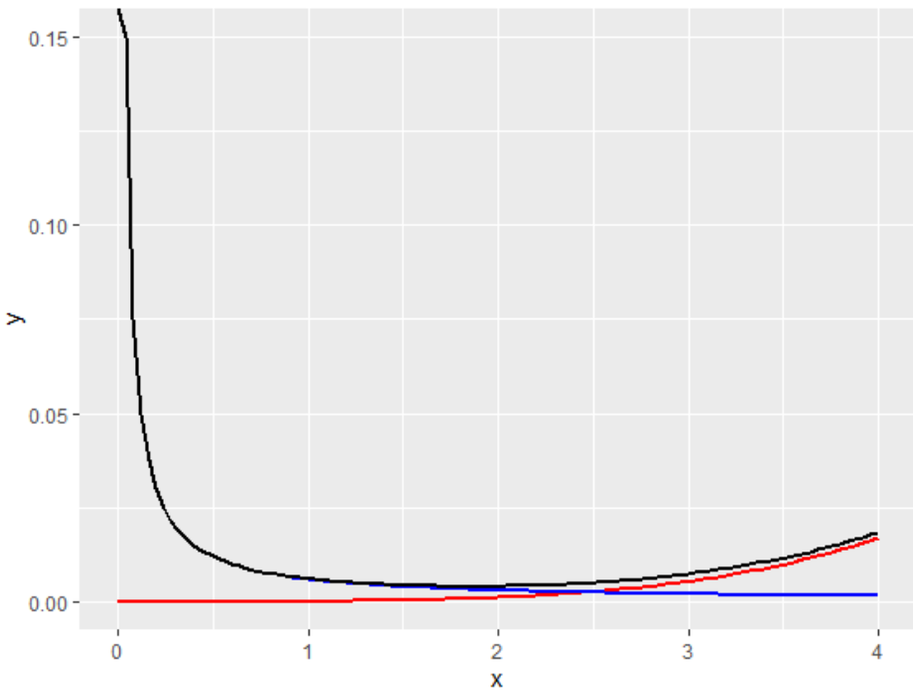


Figure 6: Plot of the Bias square term, Variance term and Total AMISE

Question 4

4.1

In order to find the optimal rule-of-thumb binwidth for the histogram pdf estimator, we minimize the AMISE of the estimated histogram pdf estimator $\hat{f}_h$. The lecture notes of lecture 3 yield the following AMISE:

$$AMISE\{\hat{f}_h\} = \frac{1}{nh} + \frac{h^2}{12}\|f'\|_2^2$$

In our derivation we minimize $AMISE\{\hat{f}_h\}$ with respect to h as

$$\frac{\delta AMISE}{\delta h} = -\frac{1}{nh^2} + \frac{1}{6}h\frac{h^2}{12}\|\|f'\|\|_2^2 = 0,$$

after which the solution is given by

$$h_{opt} = (\frac{6}{n\|\|f'\|\|_2^2})^{\frac{1}{3}}.$$

Since we do not know the actual pdf f (which is the exact reason why we apply histogram pdf or kernel density estimation), yet we need $||f'||_2^2$ in determining h_{opt} as apparent from above, we choose to assume $f \sim N(0, \sigma^2)$ such that we can compute the squared Euclidean norm of the first derivative of the pdf as follows:

$$\|f'\|_2^2 = \frac{1}{4\sqrt{\pi}\sigma^3}.$$

Plugging this value into the expression for the optimal binwidth, we obtain the following outcome:

$$h_{opt} = (\frac{24\sqrt{\pi}\sigma^3}{n})^{\frac{1}{3}} \approx 3.5\sigma n^{-\frac{1}{3}}$$

Letting $n = 1500$ equal the number of data points in our data set X, and estimate σ by taking the square root of the sample variance of X, we find $h_{opt} = 0.265$.

4.2

Next, we investigate different histogram pdf estimators for varying bindwidths. Next to the optimal rule-of-thumb binwidth h_{opt} computed in Question 4.1, we estimate the histogram pdf estimator for $\frac{1}{2}h_{opt}$ and $2h_{opt}$. Plots of the estimated pdf estimators are provided in Figures 6, 7 and 8.

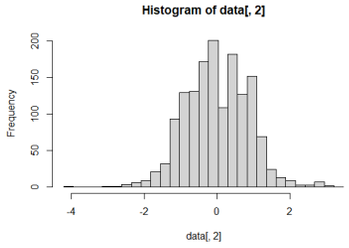


Figure 7: $h = h_{opt}$

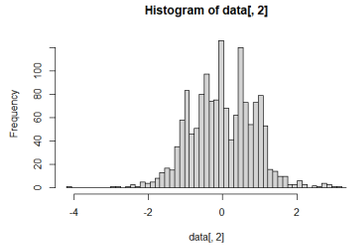


Figure 8: $h = \frac{1}{2}h_{opt}$

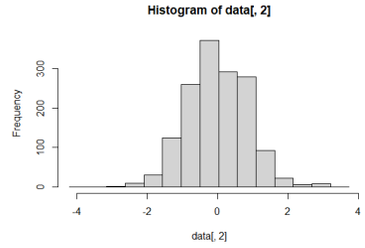


Figure 9: $h = 2h_{opt}$

The modes of a distribution are defined as the values that appear most frequently, or where a distribution takes its highest values. For $h = h_{opt} \approx 0.27$, Figure 7 shows us around 2-3 modes. For $h = \frac{1}{2}h_{opt} \approx 0.13$, Figure 8 indicates around 4-5 modes. For $h = 2h_{opt} \approx 0.53$, Figure 9 shows just 1 mode. We therefore guess the amount of modes equals approximately 3.

4.3

In order to find the Silverman's rule-of-thumb bandwidth for the kernel density estimator, we minimize the AMISE of the estimated kernel density estimator $\hat{f}_h$. The lecture notes of lecture 4 yield the following AMISE:

$$AMISE\{\hat{f}_h\} = \frac{h^4}{4}\mu_2(K^2)^4\|f''\|_2^2 + \frac{1}{nh}\|K\|_2^2$$

In our derivation we minimize $AMISE\{\hat{f}_h\}$ with respect to h, after which we obtain the h_{opt} that optimizes $AMISE\{\hat{f}_h\}$ as:

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

Again we do not know the actual pdf f even though we need $||f''||_2^2$ in determining h_{opt}. Hence, we choose to assume $f \sim N(0, \sigma^2)$ with $\mu = 0$ and $\sigma^2 = 1$ such that we can compute the squared Euclidean norm of the second derivative of the pdf as follows:

$$\|f''\|_2^2 = \frac{3}{8\sqrt{\pi}}\sigma^{-5}$$

The values of $\mu_2(K^2)$ and $\|K\|_2^2$ are computed following our choice for K, which we choose to be the Gaussian kernel. Then,

$$h_{Silverman} \approx 1.06\hat{\sigma}n^{-\frac{1}{5}}$$

Letting $n = 1500$ equal the number of data points in our data set X, and estimate σ by taking the square root of the sample variance of X, we find $h_{Silverman} = 0.2125$.

4.4

The estimated density using a Gaussian kernel function and bandwidth $h_{Silverman} = 0.2125$ computed in Question 4.3 is displayed in Figure 10. We observe two modes, one just below 0 around $x = -0.2$, and one above 0 around $x = 0.5$. We therefore conclude two modes.

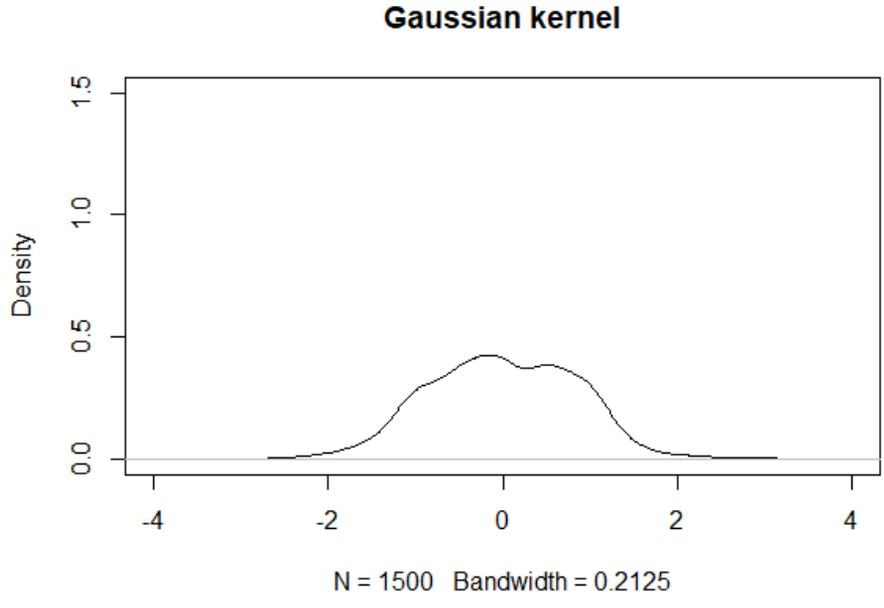


Figure 10: Estimated density

4.5

In case we want to compute the optimal bandwidth using the leave-one-out cross-validation procedure, we minimize the Integrated Squared Error (ISE) of the estimated kernel density estimator $\hat{f}_h$ given in lecture 4:

$$ISE(\hat{f}_h) = \int (\hat{f}_h(x) - f(x))^2 dx$$

which is not dependent on an assumption for the distribution function f unlike Silverman's rule-of-thumb bandwidth. We choose h to minimize the ISE, that is, we choose bandwidth $\hat{h}_{cv}$ that minimizes

$$ISE\{\hat{f}_h\} - \int f^2(x)dx \approx \int \hat{f}_h(x)dx - 2\frac{1}{n}\sum_{i=1}^{n} \hat{f_{n_1-i}}(x_i).$$

This boils down to minimizing the following expression:

$$CV(h) = \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n} K * K\left(\frac{X_j - X_i}{h}\right) - \frac{2}{n(n-1)}\sum_{i}\sum_{j\neq i}^{n} K_h(X_i - X_j)$$

where K is the density function of a $N(0,1)$ random variable. Using convolution analysis, $K * K$ is the pdf of the sum of 2 independent $N(0,1)$ variables, hence $K * K$ is the pdf of $N(0, \sqrt{2}^2$ such that:

$$K * K(u) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{u^2}{2*2}}.$$

Plugging this into the expression for $CV(h)$, we obtain the final minimization problem as

$$CV(h) = \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{4}\left(-\frac{X_j - X_i}{h}\right)^2} - \frac{2}{n(n-1)}\sum_{i}\sum_{j\neq i=1}^{n} \frac{1}{h\sqrt{2\pi}} e^{-\frac{1}{2}\left(-\frac{X_j - X_i}{h}\right)^2},$$

which we can not optimize analytically but only numerically. In specific, we determine a range of values for h, compute the values of $CV(h)$, and choose the h for which $CV(h)$ takes the lowest value. We firstly compute the $CV(h)$ scores for $h = [0.02, ..., 0.60]$ with step size 0.02, which are displayed in Figure 11.

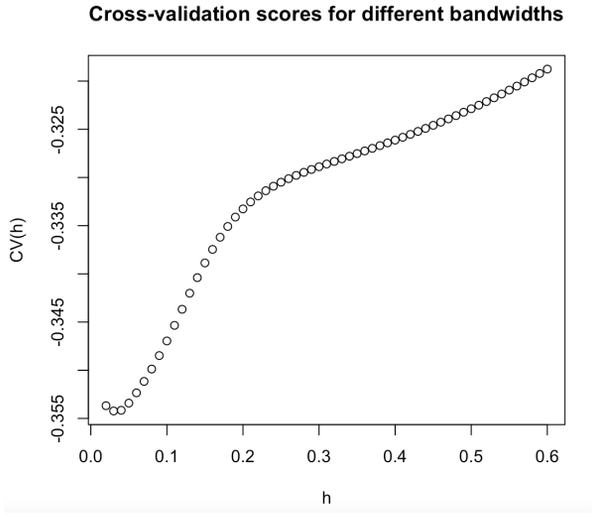


Figure 11: Values of CV(h) for h from 0.02 till 0.60 with step size 0.02.

It can be concluded that h_{CV} takes a small value, such that we repeat the cross-validation procedure for a more concentrated range of smaller values. We compute the $CV(h)$ scores for

$h = [0.01, ..., 0.11]$ with step size 0.005, which are displayed in Figure 12. This plot yields a minimal $CV(h)$ at $h_{CV} = 0.035$. In case we repeat the procedure for this range with step size 0.01, the optimal bandwidth appears to be $h_{CV} = 0.03$.

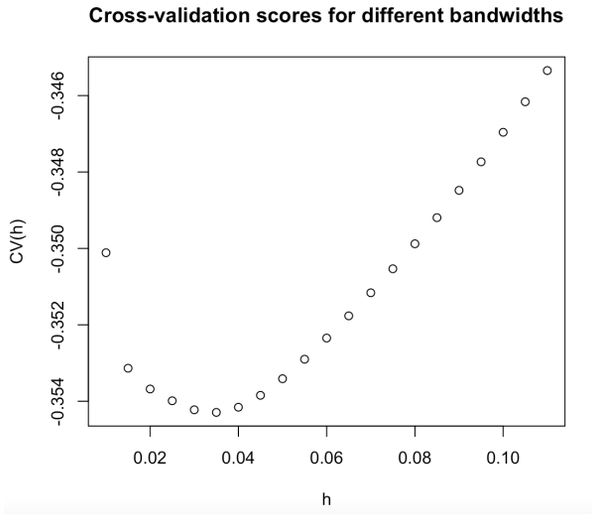


Figure 12: Values of CV(h) for h from 0.01 till 0.11 with step size 0.005.

4.6

The estimated density using a Gaussian kernel function and bandwidth $h_{CV} = 0.035$ computed in Question 4.5 is displayed in Figure 13. From the plot we observe around 4 to 5 modes, with large peaks around $x = 0$ and $x = 0.6$, and relatively lower peaks around $x = -1$, $x = -0.4$ and $x = 1$. Hence, the precise amount of modes is debatable: including all these peaks leads to 5 modes, whereas including the largest peaks of similar height amount to 2 modes.

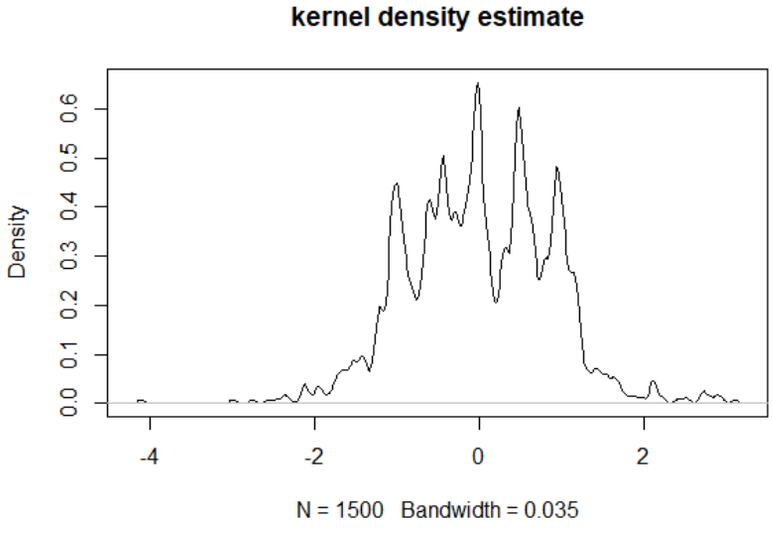


Figure 13: Estimated density with cross validated band-width

4.7

In the derivation of Silverman's rule of thumb bandwidth, we make the assumption of a normal distribution on the true pdf f, namely $f \sim N(0,1)$, to compute $\|f''\|_2^2$. This led to $h_{Silverman} = 0.2125$ with its estimated density plotted in Figure 10. This estimated pdf is smoother, less spiky and less varying than the estimated pdf based on the leave-one-out cross-validation bandwidth, which chooses h to minimize the ISE where only the Gaussian kernel density is used and no assumption on the true pdf f is made. This led to $h_{CV} = 0.035$ with its estimated density plotted in Figure 13.

Because we observed clearly irregular bulks in the given data, its shape appears differently from a normal distribution, so the assumption of the Silverman's rule of thumb seems invalid. The Silverman's rule-of-thumb bandwidth is indeed less accurate, since we expected to have two modes or more of the distribution based on the histograms as drawn in Question 4.2. As the density estimation with the cross-validated bandwidth does not make assumptions on the true pdf and shows more than two modes as expected, we consider it more trustworthy.

Question 5

By definition, the estimated pdf using kernel density estimation is given by:

$$\hat{f_h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K(x - X_i).$$

When we use the uniform kernel, this amounts to:

$$\hat{f_h}(x) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{2} I\left(\left|\frac{x - X_i}{h}\right| \leq 1\right).$$

Let B_j denote bins $j \in \mathbb{Z}$ with binwidth h. For x contained in B_j for an arbitrary j, the histogram pdf estimator is defined as follows:

$$\hat{f_h}(x) = f_j = \frac{\sum_{i=1}^{n} I(X_i \in B_j)}{nh}$$

where $I(X_i \in B_j)$ is an indicator function that equals 1 when $X_i \in B_j$, and 0 otherwise. We can rewrite this expression as follows:

$$f_h(x) = \frac{1}{n}\hat{\sum_{i=1}^{n}} K_x(X_i)$$

where $K_x(X_i) = \frac{I(X_i \in B_z)}{h}$. This kernel takes a value $\frac{1}{h}$ in case X_i is contained in bin B_x, and zero elsewhere, where h denotes the binwidth. This means that this kernel $K_x(X_i)$ corresponds to the uniform kernel with a bandwith h.

We therefore conclude that the two estimation methods are not generally the same. Particularly, the histogram pdf estimator is a specific case of the kernel density estimator.