Sunniva Olsrud Punsvik
(4130339)

ST816 Computational Statistics

PROJECT 1

April, 2025
University of Southern Denmark

Contents

1 Simulating observations from Student's *t*-distribution **2**

2 Density estimation **5**

1 Simulating observations from Student's *t*-distribution

Let X be a random variable with Student's t-distribution and parameter $\nu \in (0, \infty)$. The density function of Student's t-distribution with parameter $\nu \in (0, \infty)$ is given by:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

for $x \in \mathbb{R}$, where Γ is the gamma function.

(a) Suppose that $Y = |X|$. Derive the density function of Y.

Y is defined by the random variable X. The cumulative distribution function of Y, for $y \geq 0$:

$$F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y)$$

Since X is a random variable with Student's t-distribution, the density is symmetric around 0 (since $f_X(-x) = f_X(x)$), the cumulative density function can therefore be expressed as $F_X(-y) = 1 - F_X(y)$ and thus:

$$F_Y(y) = F_X(y) - (1 - F_X(y)) = 2F_X(y) - 1$$

To find the density function $f_Y(y)$, one has to differentiate $F_Y(y)$ with respect to y:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = 2\frac{d}{dy} F_X(y) = 2f_X(y)$$

for $y \geq 0$.

Then by substituting the expression for $f_X(y)$:

$$f_Y(y) = 2\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}\right)$$

$$f_Y(y) = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}$$

for $y \geq 0$.

(b) Can an exponentially distributed variable be used as the instrumental random variable for the acceptance-rejection algorithm to simulate observations of Y?

To determine whether an exponentially distributed random variable can serve as an instrumental distribution for the acceptance-rejection algorithm, there must exist a constant $C > 0$ such that.

$$f_Y(y) \leq C \cdot h(y), \quad \forall y \geq 0$$

where $f_Y(y)$ is the target density and $h(y)$ is the proposed density. In this case:

- $f_Y(y) = Y = |X| = f_Y(y) = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\Gamma\left(\frac{\nu}{2}\right)}} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}$
- $h(y) = \lambda e^{-\lambda y}$

The densities have to be compared and find C such that:

$$\frac{f_Y(y)}{h(y)} = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\lambda\sqrt{\nu\pi\Gamma\left(\frac{\nu}{2}\right)}} \cdot \frac{\left(1+\frac{\nu^2}{\nu}\right)^{-(\nu+1)/2}}{e^{-\lambda y}} \leq C$$

The ratio of interest:

$$\frac{\left(1+\frac{\nu^2}{\nu}\right)^{-(\nu+1)/2}}{e^{-\lambda y}} = e^{\lambda y}\left(1+\frac{y^2}{\nu}\right)^{-(\nu+1)/2}$$

This ratio needs to be analysed to see if it is bounded for all $y \geq 0$.

- As $y \to 0$:

$$e^{\lambda y}\left(1+\frac{y^2}{\nu}\right)^{-(\nu+1)/2} \approx e^{\lambda \cdot 0}(1)^{-(\nu+1)/2} = 1$$

  i.e., the ratio approaches a finite value.

- As $y \to \infty$:
  The term $e^{\lambda y}$ will grow exponentially for large values of y, while the term $\left(1+\frac{y^2}{\nu}\right)^{-(\nu+1)/2}$ decays polynomially. Therefore:

$$e^{\lambda y}\left(1+\frac{y^2}{\nu}\right)^{-(\nu+1)/2} \to \infty$$

Since the ratio goes towards infinity as $y \to \infty$, it means that it is not bounded for any finite constant C. Thus, this means that an exponential distribution cannot serve as an instrumental distribution for the acceptance-rejection algorithm in this case, because one cannot find a constant C satisfying the required condition.

(c) Propose an instrumental density g for the acceptance-rejection algorithm that simulates observations of X when $\nu \in [1, \infty)$.

To simulate observations of X that follows a Student's t-distribution with $\nu \geq 0$ using the acceptance-rejection algorithm, the proposed instrumental density g need to satisfy:

- There exists a constant C such that: $\frac{f(x)}{g(x)} \leq C$ for all $x \in \mathbb{R}$.
- $g(x)$ is easy to sample from.

From section 1 it was seen that the exponential distribution could not be an instrumental distribution since the Student's t-distribution has a heavier tail, and therefore the $\frac{f(x)}{g(x)} \leq C$ could not be satisfied. So, proposing an instrumental distribution, one could look at distributions with heavy-tailed dominance. One special case of the target distribution can be found when $\nu = 1$, the Student's t-distribution at this value of ν is also the standard Cauchy distribution, i.e., at this value their value is the same. To see if this proposed instrumental distribution can be used, one has to look at $\frac{f(x)}{g(x)}$ for $\nu > 1$. The density for the standard Cauchy distribution is given by:

$$g(x) = \frac{1}{\pi(1+x^2)}$$

While the target distribution decays like $x^{-(\nu+1)}$, the standard Cauchy distribution decays like x^{-2}, so for values $\nu > 1$, the Student's t-distribution will decay faster than the Cauchy distribution, meaning $\frac{f(x)}{g(x)}$ will remain bounded since the standard Cauchy distribution's tails are heavier.

So, using the proposed standard Cauchy:

$$\sup_{x\in\mathbb{R}} \frac{f(x)}{g(x)}$$

yields

$$\frac{f(x)}{g(x)} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}} \cdot \frac{\pi\left(1+x^2\right)}{\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}}$$

Analysing the supremum:

- At $x = 0$:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}} \cdot \frac{\pi\left(1+x^2\right)}{\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}}\pi$$

- As $x \to \infty$:

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}}\pi \cdot \frac{x^2}{(x^2/\nu)^{(\nu+1)/2}} \approx \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}}\pi \cdot \nu^{(\nu+1)/1} \cdot x^{-(\nu-1)}$$

  for $\nu > 1$ this ratio goes towards 0 as $x \to \infty$ and for $\nu = 1$ the ratio is constant at π.

From this one can see that the supremum is attained when $x = 0$ and the ratio is bounded for all $\nu \geq 1$, the ratio is bounded and therefore the requirements for acceptance-rejection algorithm are fulfilled.

Sampling from a standard Cauchy distribution is considered easy and relatively efficient since one simulates using the inverse transform method because its quantile function is known. By using the inverse transform method:

- The cumulative density function for the Cauchy: $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(\pi)$.
- The quantile function $F^{-1}(u) = \tan(\pi(u-0.5))$, where $u \sim \text{Uniform}(0,1)$.

So to generate observations of g:

Sample U:
$U \sim \text{Uniform}(0,1)$
Compute X:
$X = \tan(\pi(U-0.5))$

The expected number of simulations until acceptance for $\nu = 2$:

$$C = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\Gamma\left(\frac{\nu}{2}\right)}}\pi = \frac{\Gamma(3/2)}{\sqrt{2}\Gamma(1)}\pi = \frac{\frac{\sqrt{\pi}}{2}}{\sqrt{2}}\pi = \frac{\pi\sqrt{\pi}}{2\sqrt{2}} = \frac{\pi^{3/2}}{2^{3/2}} \approx 2.784$$

Thus, it is expected that approximately 2.784 simulations are needed before before an observation from g is accepted.

2 Density estimation

The objective here is to fit a density function to time intervals between the starts of successive eruptions of the Old Faithful geyser in Yellowstone National Park in Wyoming, United States. The dataset is built-in in `R`. The variable `waiting` is the one of interest.
(All code used for this part can be accessed using the link found here [1])

(a) Create a histogram of the `waiting` variable.

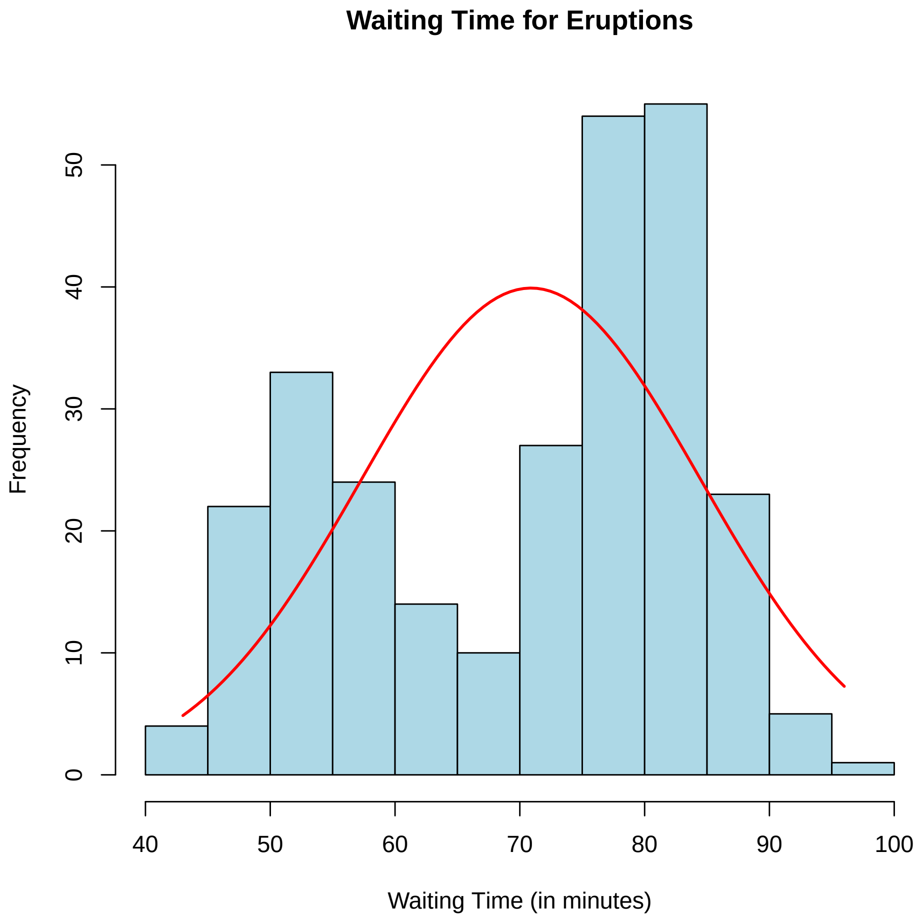

Figure 2.1: Histogram for `waiting`.

From the histogram 2.1 one can see two distinct peaks, i.e., `waiting` is bimodal. This bimodality suggests that the *Old Faithful Geyser* has two typical patterns of eruption intervals. The two peaks are also not symmetrical - the peak at around 80-85 minutes have a higher frequency compared to the peak around 50-55 minutes.

(b) Use the kernel density estimator with the standard normal kernel to estimate the density of `waiting`.

The oversmoothed bandwidth $\hat{h}_{OS}$ is given by:

$$\hat{h}_{OS} = 3\left(\frac{\hat{\sigma}^5}{35n}\right)^{1/5}$$

where $\hat{\sigma} =$ is the sample standard deviation and n is the number of observations. So:

$$\hat{h}_{OS} = 3\left(\frac{13.59497^5}{35 \cdot 272}\right)^{1/5} = 6.527881$$

Applying the 4 following bandwidths: $\hat{h}_{OS}$, $\hat{h}_{OS}/2$, $\hat{h}_{OS}/4$, and $\hat{h}_{OS}/8$ results in:

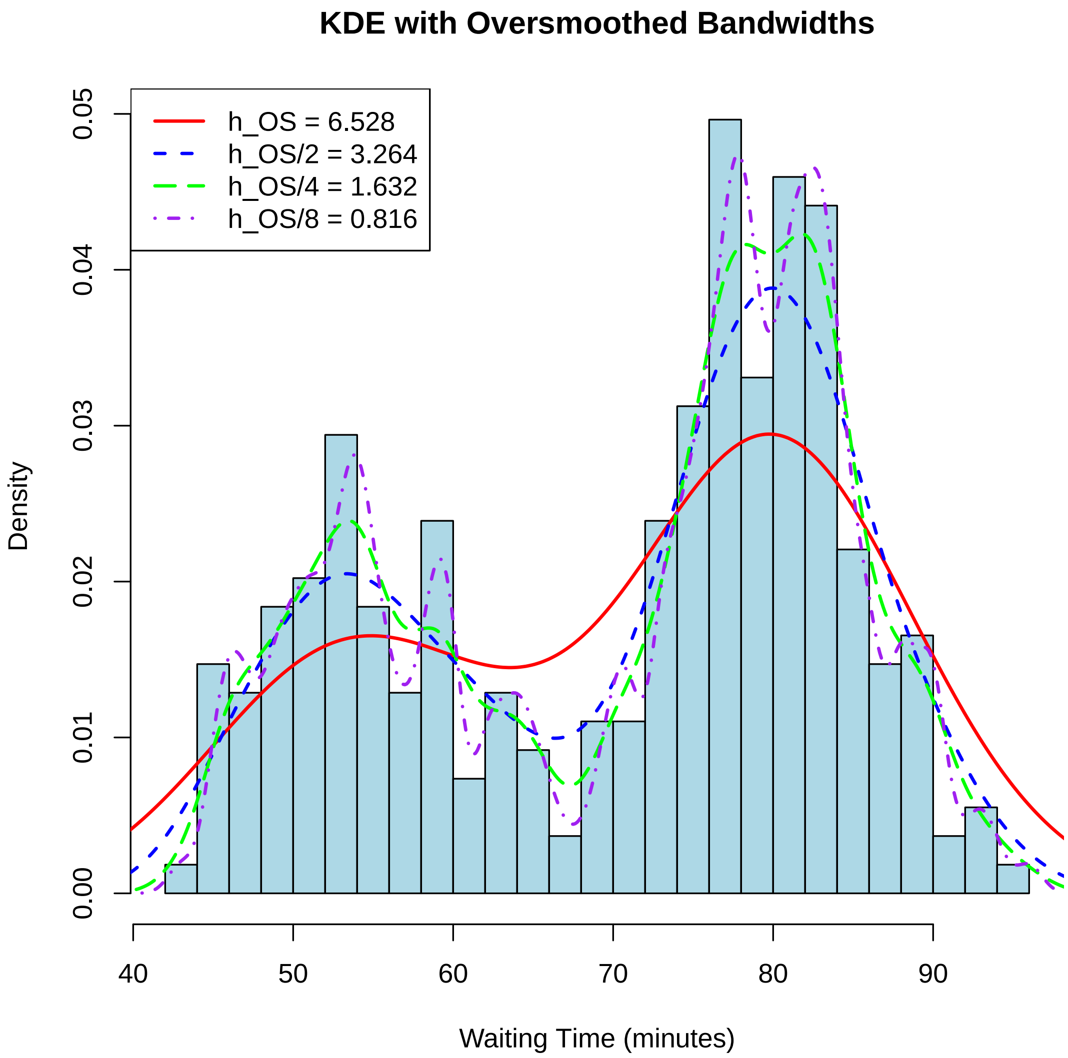

Figure 2.2: Normal kernel using different values of $\hat{h}_{OS}$.

From figure 2.2, one can comment on the following:

- $\hat{h}_{OS}$: the broad bimodial structure is somewhat visible, though most likely oversmoothed.
- $\hat{h}_{OS}/2$: There are two distinct peaks at 54 and 80 minutes.
- $\hat{h}_{OS}/4$: Again there are distinct peaks which preserves the bimodality, but begins to slightly fluctuate.
- $\hat{h}_{OS}/8$: Overfits the noise that results in spurious peaks.

From this one could conclude that using the bandwidth value of $\hat{h}_{OS}/2 = 3.264$ is the best fit for the `waiting` variable.

(c) Would the normal scale bandwidth selector for the `waiting` variable be appropriate?

Normal scale bandwidth selector assumes the underlying distribution is normally distributed and calculates the bandwidth:

$$h = 1.06 \cdot \hat{\sigma} \cdot n^{-1/5}$$

where $\hat{\sigma}$ is the sample standard deviation and n is the number of observations.

This selector is not ideal since `waiting` is bimodial and the normal scale bandwidth selector assumes normality and thus, `waiting` violates this. Additionally, there would be risks for oversmoothing due to this bimodality, blurring the true peaks. One could forcibly apply the selector, but it would underfit the data and smooth out bimodality, so no, it does not make sense to use the normal scale bandwidth selector for the `waiting` variable.

(d) Calculate the bandwidths using the following selectors for the standard normal kernel and report their values: normal scale, pseudo-likelihood, least squares cross-validation, biased cross-validation, and Sheather-Jones.

Table 1: Summary of bandwidths.

Selector	Normal Scale (hNS)	Pseudo-Likelihood (hPL)	Least Squares CV (hUCV)	Biased CV (hBCV)	Sheather-Jones (hSJ)
Bandwith (h)	4.69	2.50	2.66	2.60	2.50

From table 1 and figure 2.3, one can see that hNS oversmooths, i.e., there are not two distinct peaks, useful for unimodal normal data. The remaining selectors are close in value and overlaps in the histogram, meaning that the remaining 4 selectors are suitable for the *Old Faithful Geyser* data set. However, if one had to choose one selector, one ought to look at the differences in smoothing and noise. hPL, hUCV, hBCV, and hSJ all preserve bimodality, from a theoretical and practical perspective, hUCV is known to undersmooth (which is fine for exploratory analysis) while hBCV may oversmooth (thus providing conservative estimations). hPL is good for general use despite being slightly senstitive to outliers and the hSJ, while computationally expensive, is the most reliable.

(e) Consider the kernels: standard normal, Epanechnikov, rectangular, triangular, and biweight using Sheather-Jones bandwidth selector for the standard normal kernel.

Sheather-Jones kernel is optimised for the standard normal kernel. One has to convert equivalent bandwidths for other kernels by using the canonical bandwidth scaling factor on kernel efficiency:

$$h_{\text{new}} = h_{SJ}$$

The Asymptotic Mean Integrated Squared Error (AMISE) for a kernel K is given by:

$$h_{\text{AMISE}} = \left(\frac{R(K)}{n\mu_2(K)^2 R(f'')} \right)^{1/5}$$

Where $R(K)$ is the roughness of the kernel, $\mu_2 K$ is the variance of the kernel, and $R(f'')$ is the roughness of the second derivative of the true density f.

To compare bandwidths across kernels, one has to eliminate the second derivative of the true density since it is unknown but constant for a given dataset. This can be done by taking the ratio between two kernels K_1 and K_2:

$$\frac{h_1}{h_2} = \left(\frac{R(K_1)\mu_2(K_2)^2}{R(K_2)\mu_2(K_1)^2} \right)^{1/5}$$

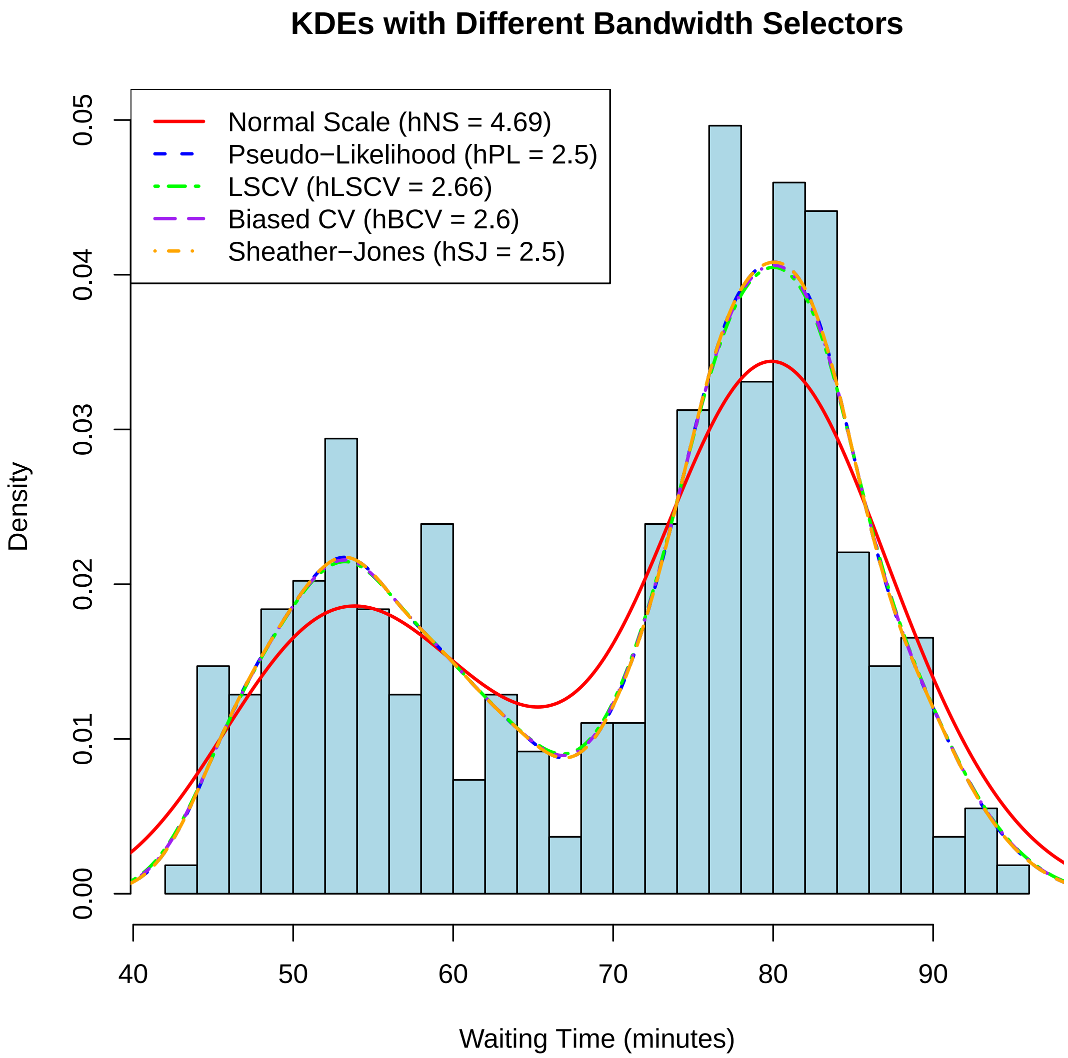

Figure 2.3: Bandwidths of different selectors.

which results in the scaling factor needed to convert h_{SJ} to another kernel.
For the standard normal kernel $R(K_1) = \frac{1}{2\sqrt{\pi}}$, and $\mu_2(K_1) = 1$

Table 2: Scaling factors and transformed h for each kernel.

Kernel	**Scaling Factor (α)**	**Scaling**	**Transformed h**
Standard Normal	1.000	h_{SJ}	2.50
Epanechnikov	1.719	$h_{SJ} \cdot 0.776$	1.94
Rectangular	1.741	$h_{SJ} \cdot 0.768$	1.92
Triangular	1.888	$h_{SJ} \cdot 0.727$	1.82
Biweight	2.036	$h_{SJ} \cdot 0.685$	1.71

From table 2 and figure 2.4 all the kernels capture bimodality except rectangular, which creates blocky artifacts. The Epanechnikov kernel is theoretically optimal for minimising AMISE - offering a good balance between bias and variance - and captures the *Old Faithful Geyser* dataset well. The biweight kernel also captures the bimodal structure, though with smoother tails. The standard normal kernel is a safe choice but oversmooths. The triangular kernel have the same blocky artifacts like the rectangular,

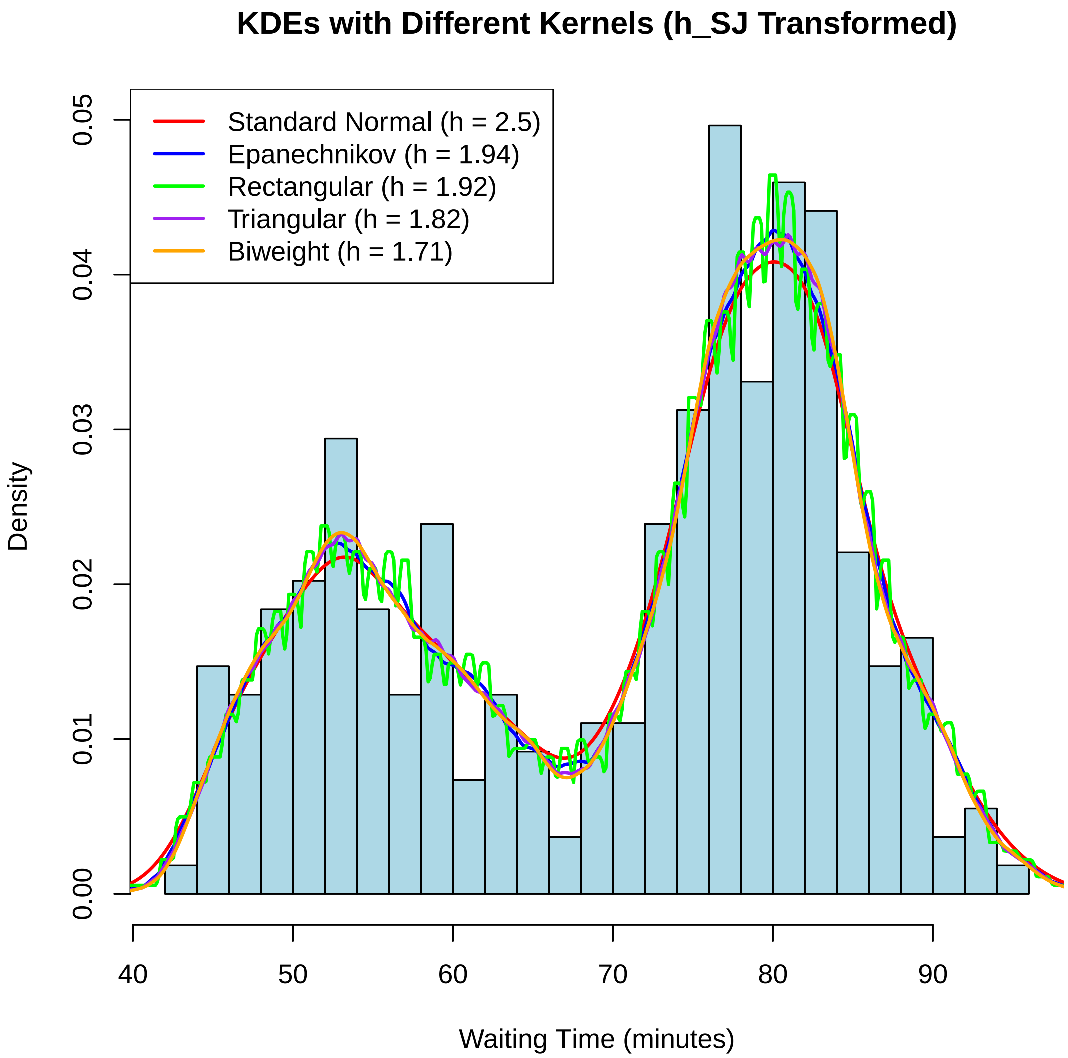

Figure 2.4: Bandwidths of different kernels.

but the triangular is smoother.

Thus, the Epanechnikov kernel is a reasonable choice for the waiting variable, but the biweight kernel could aslo be used.

References

[1] https://github.com/sunnivaolsrud/ST816-projects.git.