Assignment 2

Shaaz

2025-02-07

```
personal_number <- 180
```

Question 1

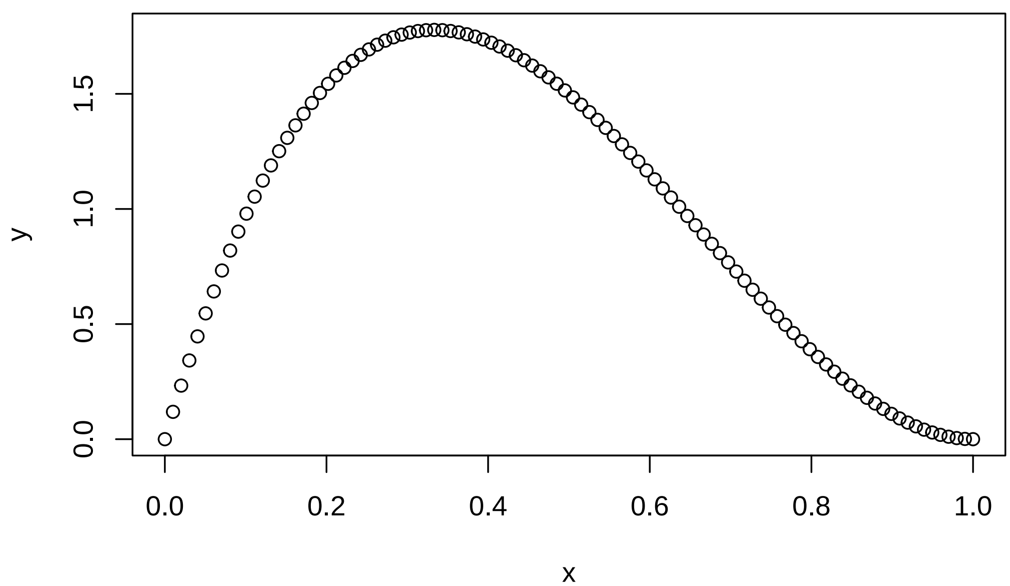
a) Mathematical Analysis

$$
\begin{aligned}
B &= \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \\
&= \frac{(1)(2)}{(24)} = \frac{1}{12}
\end{aligned}
$$

$$
\begin{aligned}
\therefore f(x) &= \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \\
&= \frac{x(1-x)^2}{(1/12)} \\
&= 12x(1-x)^2
\end{aligned}
$$

When plotting this density, we get:

```
x <- seq(0, 1, length.out = 100)
y <- 12 * x * (1 - x)^2
plot(x,y)
```



```
max(y)
```

```
## [1] 1.777778
```

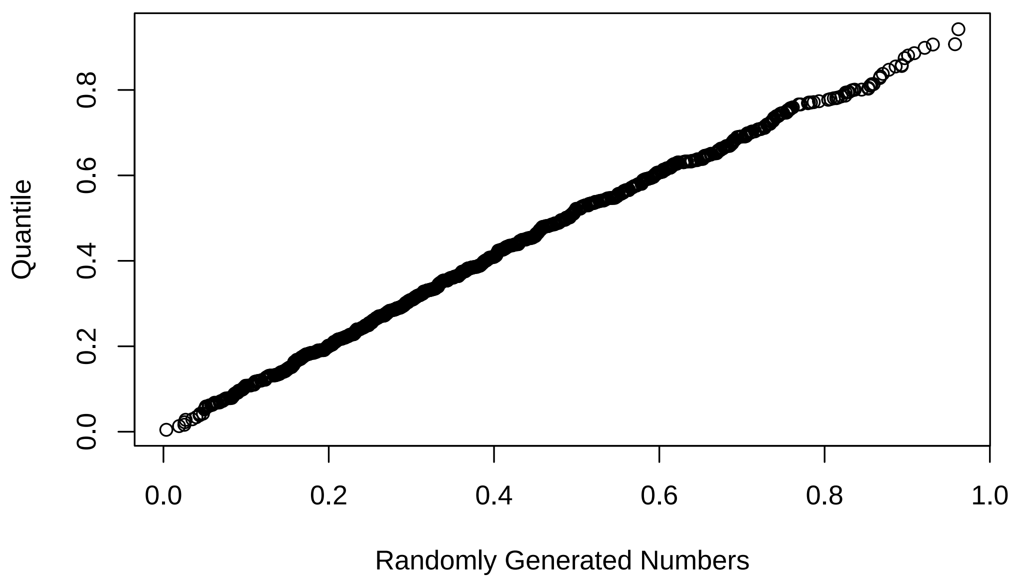
And so, we can now use the acceptance/rejection method, with a = 0, b = 1, c = 1.78

b) R code generating 1000 numbers

```
iterations <- 1000
random <- array(NA, dim=c(1,iterations))
set.seed(personal_number)
while (iterations > 0) {
  U = runif(1, 0, 1)
  V = runif(1, 0, 1.78)
  if (12 * U * (1 - U)^2 >= V) {
    random[1, iterations] <- U
    iterations <- iterations - 1
  }
}
```

d) qq-Plot of the 1000 numbers

```
set.seed(personal_number)
r <- qbeta(runif(1000), 2, 3)
qqplot(random, r, xlab="Randomly Generated Numbers", ylab ="Quantile")
```

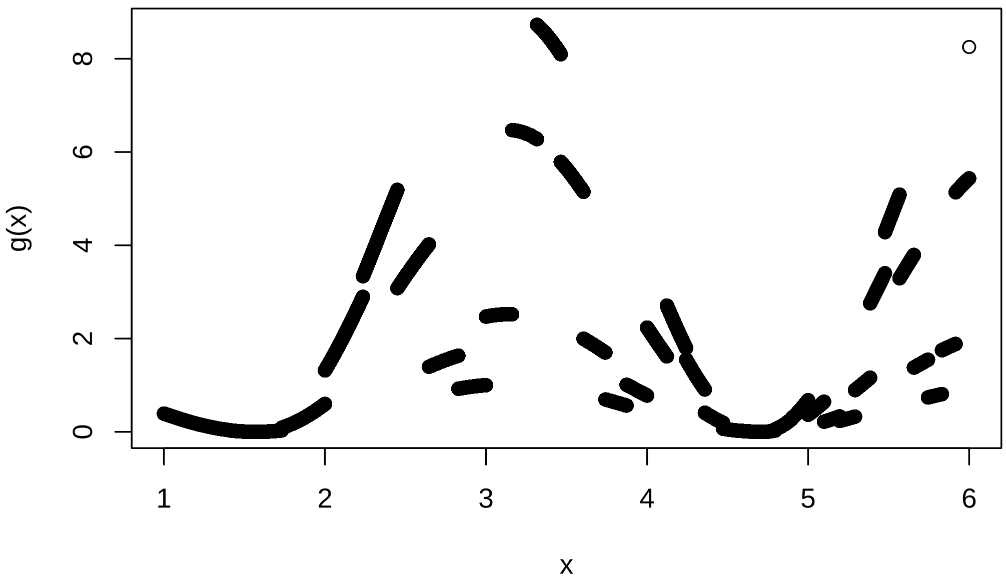


Hence, our randomly generated numbers are good.

Question 2

a) Plotting g(x)

```r
g <- function(x) {
  return((cos(x) * (2 - sin(floor(x^2))))^2)
}
x <- seq(1, 6, length.out=200000)
plot(x, g(x))
```




```r
mean(g(x))
```

```
## [1] 1.86658
```

b) Finding minimal N

$$N \geq \frac{1}{4\delta\epsilon^2} = \frac{1}{(4)(0.05)(0.005)^2}$$
$$\therefore N \geq 200\,000$$

c) Computing integral and variance with minimal N

$$\int_0^6 g(x) \approx (v-u)\frac{1}{N}\sum_{i=1}^{N} g(X_i)$$
$$\approx (6-0)(1.86658) = 11.19948$$
$$\sigma^2 \approx \frac{1}{N-1}\sum_{i=1}^{N}(g(X_i) - S_N)^2$$

```r
(mean((g(x) - mean(g(x)))^2)*200000)/(199999)
```

```
## [1] 4.345713
```

$$\therefore \sigma^2 = 4.345713$$

d) Recalculating minimal N

$$\begin{aligned} N &= \frac{\sigma^2}{\delta\epsilon^2} = \frac{4.345713}{(0.05)(0.005)^2} \\ N &= 3\,476\,570 \end{aligned}$$

e) Recalculating integral

```r
set.seed(personal_number)
x <- seq(1, 6, length.out=3476570)
mean(g(x))
```

```
## [1] 1.866566
```

$$\begin{aligned} \int_0^6 g(x) &\approx (v-u)\frac{1}{N}\sum_{i=1}^{N} g(X_i) \\ &\approx (6-0)(1.866566) = 11.199396 \end{aligned}$$

As we can see, we are getting a very similar, but more accurate result as previously.

Question 3

a) Extracting bandwidth

```r
set.seed(personal_number)
library(MASS)
d <- density(geyser$waiting)
attributes(d)
```

```
## $names
## [1] "x"         "y"         "bw"        "n"         "old.coords"
## [6] "call"      "data.name" "has.na"
##
## $class
## [1] "density"
```

```r
d$bw
```

```
## [1] 3.997796
```

b) Bias of bandwidth estimate

```
set.seed(personal_number)
density(sample(geyser$waiting, 10, replace=TRUE))$bw
```

```
## [1] 7.1802
```

```
set.seed(personal_number)
density(sample(geyser$waiting, 100, replace=TRUE))$bw
```

```
## [1] 5.236356
```

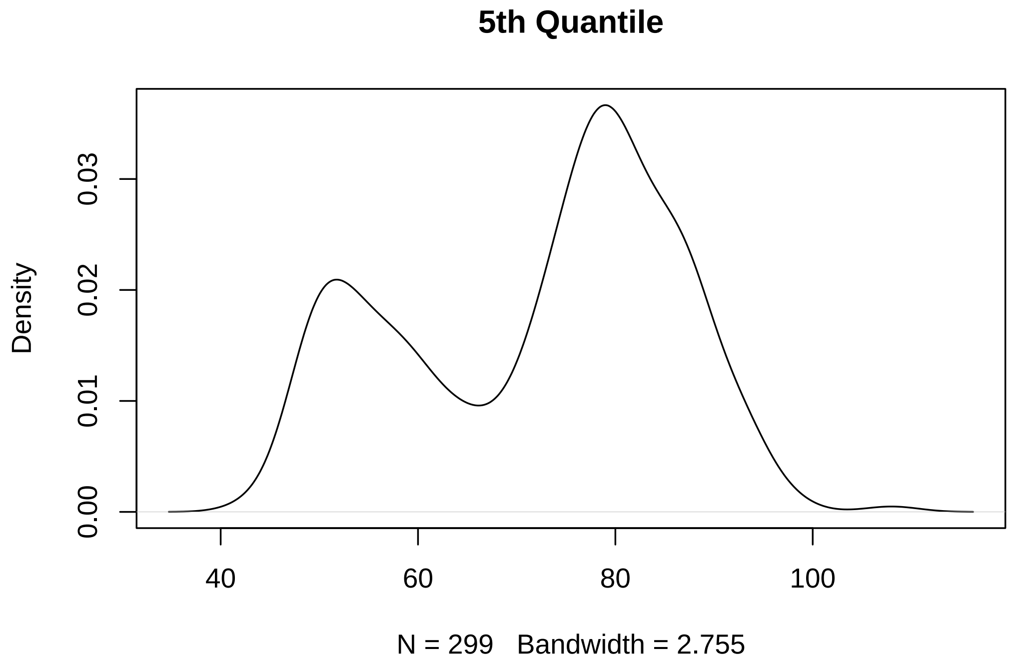
```
set.seed(personal_number)
density(sample(geyser$waiting, 1000, replace=TRUE))$bw
```

```
## [1] 3.158051
```

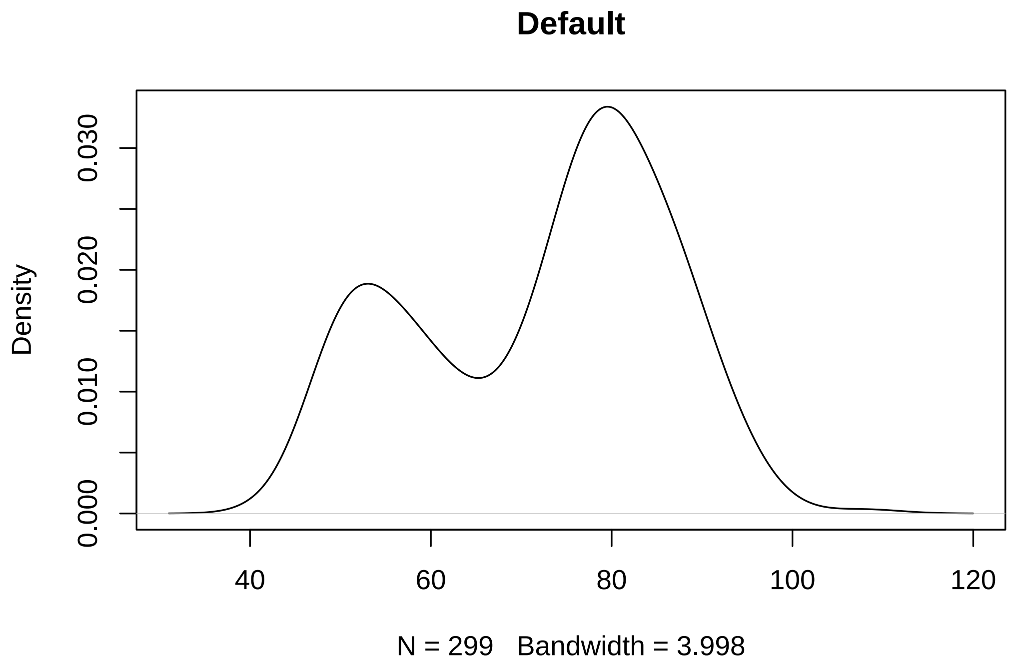
Based on the results above, I believe that the bandwidth estimates are biased.

c) Quantile plots

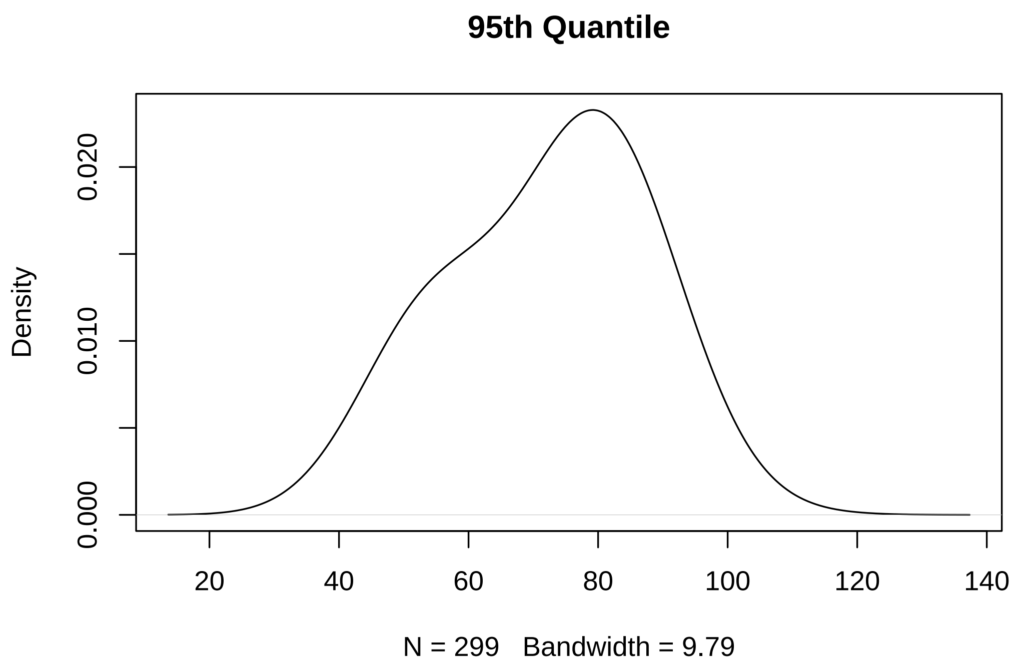
```
set.seed(personal_number)
bw_samples <- numeric(2000)
for (i in 1:2000) {
  bw_samples[i] <- density(sample(geyser$waiting, 10, replace=TRUE))$bw
}
plot(density(geyser$waiting, bw = quantile(bw_samples, 0.05)), main = "5th Quantile")
```



```
plot(density(geyser$waiting), main = "Default")
```




```
plot(density(geyser$waiting, bw = quantile(bw_samples, 0.95)), main = "95th Quantile")
```



We notice here that the 5th and default bandwidth are about the same, whereas the 95th is much more smooth

Question 4

$$
\begin{aligned}
\bar{Y}\left(1-\bar{Y}\right) &= \bar{Y}\left(1-2\bar{Y}+\bar{Y}\right) \\
&= \bar{Y}-2\bar{Y}\bar{Y}+\bar{Y}^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i - 2\bar{Y}\frac{1}{N}\sum_{i=1}^{N} Y_i + \frac{N}{N}\bar{Y}^2 \\
&= \frac{1}{N}\left(\sum_{i=1}^{N} Y_i - 2\bar{Y}\sum_{i=1}^{N} Y_i + N\bar{Y}^2\right) \\
&= \frac{1}{N}\left(\sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} 2\bar{Y}Y_i + \sum_{i=1}^{N} \bar{Y}^2\right), \text{ by linearity of summation} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i - 2\bar{Y}Y_i + \bar{Y}^2, \text{ by linearity of summation} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2, \text{ since } Y_i \in \{0,1\}, \text{ then } Y_i = Y_i^2 \\
\bar{Y}\left(1-\bar{Y}\right) &= \frac{1}{N}\sum_{i=1}^{N} \left(Y_i-\bar{Y}\right)^2
\end{aligned}
$$