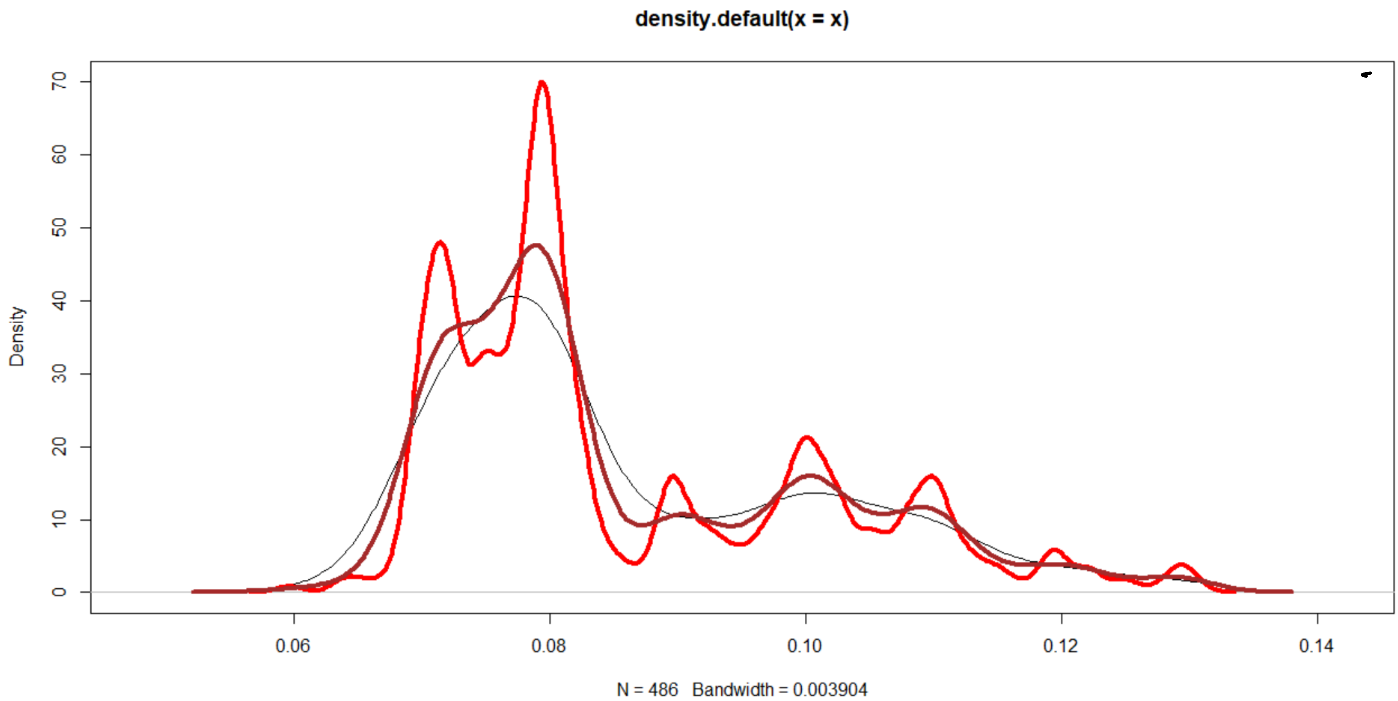
Q1

a).

```
#Q1 a
x<-scan("stamp.txt")
plot(density(x),ylim=c(0,70))
lines(density(x,bw=0.001),col="red",lwd=4)#7modes
lines(density(x,bw=0.0025),col="brown",lwd=4)#5modes
```



The bandwidth parameter h controls the amount of smoothing:

As h increases, the estimator $\hat{f}_n(x)$ becomes smoother.

The function density scale its kernels to have variance 1 (as well as mean 0)

When bendwidth = 0.001 we will have 7 modes.

When bendwidth = 0.0025 we will have 5 modes.

b) $\mathcal{L}(h)=\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1}{nh}\sum_{j=1}^{n}w\left(\frac{X_i-X_j}{h}\right)\right)=\frac{1}{n}\sum_{i=1}^{n}\ln\left(\hat{f}_h(X_i)\right).$

Assume that $W(0)>0$ & for $x\neq 0$, $h^{-1}w(x/h)\to 0$ as $h\downarrow 0$

i). WTS: $\mathcal{L}(h)\uparrow\infty$ as $h\downarrow 0$

$$h^{-1}W\left(\frac{x}{h}\right)=\frac{1}{h}W\left(\frac{x}{h}\right)$$

WTS: $\frac{1}{nh}\sum_{j=1}^{n}W\left(\frac{x_i-x_j}{h}\right)\to\infty$

$$\frac{1}{nh}\sum_{j=1}^{n}W\left(\frac{x_i-x_j}{h}\right)=\frac{1}{nh}\sum_{x_j\neq x_i}^{n}W\left(\frac{x_i-x_j}{h}\right)+\frac{1}{nh}\sum_{x_j=x_i}^{n}W\left(\frac{0}{h}\right)$$

① as $h\downarrow 0$ $\frac{1}{nh}\sum_{x_j\neq x_i}^{n}W\left(\frac{x_i-x_j}{h}\right)\to 0$ since

for $x\neq 0$, $h^{-1}w(x/h)\to 0$ as $h\downarrow 0$

② as $h\downarrow 0$ $\frac{1}{nh}\underbrace{\sum_{x_j=x_i}^{n}W(0)}_{>0\text{ (constant)}}\to\infty$

Thus, $\frac{1}{nh}\sum_{j=1}^{n}W\left(\frac{x_i-x_j}{h}\right)\to 0+\infty=\infty$

Later, as $h\downarrow 0$

$$\mathcal{L}(h)=\frac{1}{n}\sum_{i=1}^{n}\underbrace{\ln\underbrace{\left(\frac{1}{nh}\sum_{j=1}^{n}W\left(\frac{x_i-x_j}{h}\right)\right)}_{\to\infty}}_{\to\infty}\to\infty$$

ii) WTS: $CV(h) \to -\infty$ as $h \downarrow 0$ and $h \uparrow \infty$

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{(n-1)h}\sum_{j\neq i} w\left(\frac{X_i - X_j}{h}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\widehat{f}_h^{(-i)}(X_i)\right)$$

where

$$\widehat{f}_h^{(-i)}(x) = \frac{1}{(n-1)h}\sum_{j\neq i} w\left(\frac{x - X_j}{h}\right)$$

WTS: $\frac{1}{(n-1)h} \sum_{j \neq i} w\left(\frac{x_i - x_j}{h}\right) \to 0$ as $h \downarrow 0$ and $h \uparrow \infty$

For $\frac{1}{(n-1)h} \sum_{j \neq i} w\left(\frac{x_i - x_j}{h}\right)$ we know that $x_i - x_j \neq 0$

When $h \uparrow \infty$ $\frac{1}{(n-1)h} \sum_{j \neq i} \underbrace{w\left(\frac{x_i - x_j}{h}\right)}_{\text{real number} > 0} \to \frac{1}{(n-1)h} \sum_{j \neq i} w(0)$

$\to 0$ as $h \to \infty$

When $h \downarrow 0$ $\frac{1}{(n-1)h} \sum_{j \neq i} w\left(\frac{x_i - x_j}{h}\right) \to 0$

Since $\frac{1}{h} w\left(\frac{x}{h}\right) \to 0$ as $h \downarrow 0$

Thus, as $h \downarrow 0$ and $h \uparrow \infty$

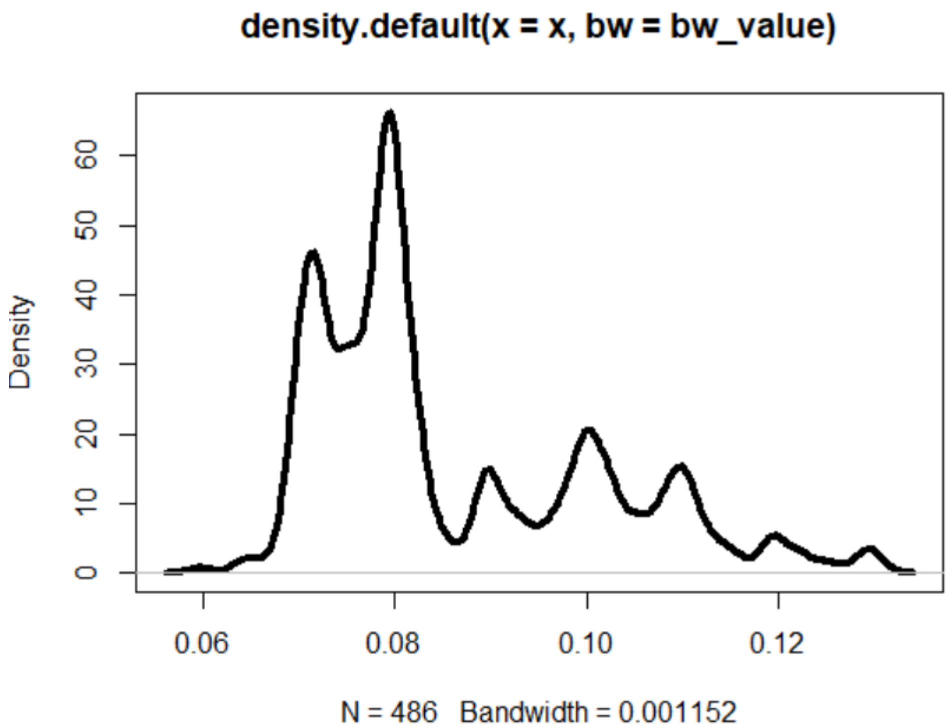
$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{1}{(n-1)h} \sum_{j \neq i} w\left(\frac{x_i - x_j}{h}\right)\right)$$

$$\to \frac{1}{n} \sum_{i=1}^{n} \ln(0)$$

$$CV(h) \to -\infty.$$

c)

```
> #Q1 c
> kde.cv <- function(x,h) {
+   n <- length(x)
+   if (missing(h)) {
+     r <- density(x)
+     h <- r$bw/10 + 3.9*c(0:100)*r$bw/100
+   }
+   cv <- NULL
+   for (j in h) {
+     cvj <- 0
+     for (i in 1:n) {
+       z <- dnorm(x[i]-x,0,sd=j)/(n-1)
+       cvj <- cvj + log(sum(z[-i]))
+     }
+     cv <- c(cv,cvj/n)
+   }
+   r <- list(bw=h,cv=cv)
+   r
+ }
>
> r <- kde.cv(x)
> plot(r$bw,r$cv) # # plot of bandwidth versus CV
> bw_value <-r$bw[r$cv==max(r$cv)] # bandwidth maximizing CV
>
> plot(density(x,bw=bw_value),lwd=4)
```



This density have 7 modes.

Q2.

a) $\mathcal{L}_F(t) = \frac{1}{\mu(F)} \int_0^t F^{-1}(s)\, ds$ with $\mu(F) = \int_0^1 F^{-1}(s)\, ds$.

WTS $\mathcal{L}'_F(MPS(F)) = 1$

$$\mathcal{L}'_F(MPS(F)) = \frac{1}{\mu(F)} F^{-1}(MPS(F))$$

$$= \frac{1}{\mu(F)} F^{-1}(F(\mu(F)))$$

$$= \frac{1}{\mu(F)} \cdot \mu(F)$$

$$= 1$$

b) Assume $\mathcal{L}_F(t) = t^{\alpha+1}$ for some $\alpha \geq 0$

$\mathcal{L}_F(MPS(F)) = [MPS(F)]^{\alpha+1}$

$\mathcal{L}'_F(MPS(F)) = (\alpha+1) MPS(F)^{\alpha} = 1$ by a).

$MPS(F) = \left(\frac{1}{\alpha+1}\right)^{\frac{1}{\alpha}}$

$$MIS(F) = \mathcal{L}_F(MPS(F)) = \left(\left(\frac{1}{\alpha+1}\right)^{\frac{1}{\alpha}}\right)^{\alpha+1} = \left(\frac{1}{\alpha+1}\right)^{1+\frac{1}{\alpha}}$$

c)

```
> y<-scan("incomes.txt")
Read 200 items
> MPS=sum(y<mean(y))/length(y)
> MPS
[1] 0.69
>
> loo <- NULL
> for (i in 1:200) {
+  yi <- y[-i]
+ loo <- c(loo,sum(yi<mean(yi))/length(yi))
+ }
>
> jackse <- sqrt(199*sum((loo-mean(loo))^2)/200)
> jackse
[1] 0.07811778
```

The estimate of MPS(F) is 0.69 and use the jackknife to give an estimate its standard error is 0.0781.

d). $\hat{L}_F(t) = \frac{1}{n\bar{x}} \sum_{i=1}^{[nt]} x_{(i)}$

$L_F(MPS(F)) = MIS(F)$ Thus, $\hat{L}_F(t) = \widehat{MIS}(F)$

$\widehat{MPS}(F) = \frac{1}{n} \sum_{i=1}^{n} I(x_i < \bar{x})$

$\hat{L}_F(t) = \frac{1}{n\bar{x}} \sum_{i=1}^{[n \cdot \frac{1}{n} \sum_{i=1}^{n} I(x_i < \bar{x})]} x_{(i)} = \frac{1}{n\bar{x}} \sum_{i=1}^{\sum_{i=1}^{n} I(x_i < \bar{x})} x_{(i)}$

$= \frac{1}{n\bar{x}} [x_{(1)} + \cdots + x_{(\sum_{i=1}^{n} I(x_i < \bar{x}))}]$

```
> #Q2-d
> MIS=1/(length(y)*mean(y))*sum(sort(y)[1:sum(y<mean(y))])
> MIS
[1] 0.3480396
>
> loo <- NULL
> for (i in 1:200)
+   {
+   yi <- y[-i]
+   loo <- c(loo,1/(length(yi)*mean(yi))*sum(sort(yi)[1:sum(yi<mean(yi))]))
+ }
>
> jackse2 <- sqrt(199*sum((loo-mean(loo))^2)/200)
> jackse2
[1] 0.08170359
>
```

The estimate of MIS(F) is 0.348 and use the jackknife to estimate its standard error is 0.08170359.