1) a) For 3 nodes, bandwidth needs to be about 5

For 2 nodes, bandwidth needs to be about 4

1) b)

For (i) and (ii) below, assume that $w(0) > 0$ and for $x \neq 0$, $h^{-1}w(x/h) \to 0$ as $h \downarrow 0$ (which is true for most commonly used kernels).

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

(ii) In the case where $X_1, \cdots, X_n$ are distinct (that is, no tied observations), show that $\mathrm{CV}(h) \to -\infty$ as $h \downarrow 0$ and $h \uparrow \infty$.

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

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

↳ splitting summation into two summation.

Then taking limit as $h \to 0$:

$$\lim_{h\to 0} L(h) = \lim_{h\to 0} \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{nh}\sum_{j\neq i}^{n} w\left(\frac{x_i - x_j}{h}\right) + \frac{1}{nh}\sum_{j=i} w\left(\frac{x_i - x_j}{h}\right)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln\left(\lim_{h\to 0} \frac{1}{n}\sum_{j\neq i}^{n} \frac{1}{h} w\left(\frac{x_i - x_j}{h}\right) + \frac{1}{n}\sum_{j=i} \lim_{h\to 0} \frac{1}{h} w(0)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{n}\sum_{j\neq i}^{n} \lim_{h\to 0} \frac{1}{h} w\left(\frac{x_i - x_j}{h}\right) + \frac{1}{n}\sum_{j=i} \lim_{h\to 0} \frac{1}{h} w(0)\right)$$

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

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

$$= +\infty$$

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

Taking lim as $h \to 0$

$$\lim_{h\to 0} CV(h) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\lim_{h\to 0} \frac{1}{(n-1)}\sum_{j\neq i} \frac{1}{h} w\left(\frac{x_i - x_j}{h}\right)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{(n-1)}\sum_{j\neq i} \lim_{h\to 0} \frac{1}{h} w\left(\frac{x_i - x_j}{h}\right)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{(n-1)}\sum 0\right) \qquad \text{By assumption}$$

$$= \frac{1}{n}\sum_{i=1}^{n} -\infty$$

$$= -\infty$$

→ Cont'd

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

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

$$= \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{1}{(n-1)}\sum_{j\neq i} 0\right) \qquad \text{By assumption}$$

$$= \frac{1}{n}\sum_{i=1}^{n} -\infty$$

$$= -\infty$$

2. Suppose that F is a distribution concentrated on the positive real line (i.e. $F(x) = 0$ for $x < 0$). If $\mu(F) = E_F(X)$ then the **mean population share** of the distribution F is defined as $\text{MPS}(F) = F(\mu(F)-) = P_F(X < \mu(F))$. (When F is a continuous distribution, $F(\mu(F)-) = F(\mu(F))$.) For most income distributions, $\text{MPS}(F) > 1/2$ with $\text{MPS}(F) = 0$ if (and only if) all incomes are equal and $\text{MPS}(F) \to 1$ as $\text{Gini}(F) \to 1$.

(a) Suppose that F is a continuous distribution function with Lorenz curve $\mathcal{L}_F(t)$. Show that $\text{MPS}(F)$ satisfies the condition

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

where $\mathcal{L}'_F(t)$ is the derivative (with respect to t) of the Lorenz curve.

a) assume that F is a continuous distribution function

Then $\text{MPS}(F) = F(\mu(F) - x) = F(\mu(F))$

Since $\mathcal{L}_F(\tau) = \frac{1}{\mu(F)}\int_0^{\tau} F^{-1}(s)\,ds$

Then $\mathcal{L}_F(\tau)' = \frac{1}{\mu(F)}\left(F^{-1}(\tau) - F^{-1}(0)\right)$

So $\mathcal{L}_F(\text{MPS}(F))' = \frac{1}{\mu(F)}\left(F^{-1}(F(\mu(F))) - F^{-1}(0)\right)$

$$= \frac{1}{\mu(F)}\left(\mu(F) - 0\right) \qquad \text{Since } F^{-1}(0) = 0$$

$$= 1$$

2) (b) Suppose that F is a Weibull distribution with pdf

$$f(x;\alpha,\sigma) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right] \quad \text{for } x \geq 0$$

where $\alpha, \sigma > 0$. Compute MPS(F). (Hint: It will depend only on α.) What happens to MPS(F) as $\alpha \to 0$? As $\alpha \to \infty$?

$MPS(F) = F(\mu(F))$

$F(x) = \int_0^x f(x)\,dx$

Then, $F(x) = \frac{\alpha}{\sigma}\int_0^x \left(\frac{x}{\sigma}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right] dx$

using change of variable, let $\left(\frac{x}{\sigma}\right)^{\alpha} = y$, then $x = y^{\frac{1}{\alpha}}\sigma$

$dy = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1} dx$

Then, $F(x) = \int \exp[-y]\,dy$

so $F(x) = -\exp\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right] + c$

Then, $MPS(F) = -\exp\left[-\left(\frac{\mu(F)}{\sigma}\right)^{\alpha}\right] + c$

Using same substitution $x = y^{\frac{1}{\alpha}}\sigma$

Note : $\mu(F) = E(F(x)) = \frac{\alpha}{\sigma}\int_0^{\infty} x\left(\frac{x}{\sigma}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right] dx$

$= \int_0^{\infty} \sigma y^{\frac{1}{\alpha}} \exp(-y)\,dy$

$= \sigma\Gamma\left(\frac{1}{\alpha}+1\right)$ $\leftarrow$ gamma function

so $MPS(F) = -\exp\left[-\left(\frac{\sigma\Gamma\left(\frac{1}{\alpha}+1\right)}{\sigma}\right)^{\alpha}\right] + c$

$MPS(F) = -\exp\left[-\left(\Gamma\left(\frac{1}{\alpha}+1\right)\right)^{\alpha}\right] + c$

For lim MPS(F) as $\alpha \to 0$ and $\alpha \to \infty$, Check R code graphs

Essentially as $\alpha \to 0$, $MPS(F) \to \infty$

as $\alpha \to \infty$, $mps(F) \to 0.5614$

(d) Suppose in part (c), we are assume that the incomes come from a log-Normal distribution; in other words, the natural logarithms are Normal with unknown mean and variance μ and σ^2. Show that for the log-Normal distribution,

$$\text{MPS}(F) = \Phi(\sigma/2) = \psi(\sigma)$$

where $\Phi(x)$ is the $\mathcal{N}(0,1)$ cdf. If $\widehat{\sigma}_n$ is the MLE of σ based on n observations from the log-Normal distribution, find the limiting distribution of $\sqrt{n}(\psi(\widehat{\sigma}_n) - \psi(\sigma))$. Use this result to give an estimate of the standard error of $\psi(\widehat{\sigma})$ for the data in part (c).

Step 1: Find limiting distribution of $\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2)$

Since $\dfrac{(n-1)\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}$

But $\chi^2_{n-1} \sim N(n-1, 2(n-1))$

So $\dfrac{(n-1)\hat{\sigma}_n^2}{\sigma^2} \sim N(n-1, 2(n-1))$

Then $\text{Var}\left(\dfrac{(n-1)\hat{\sigma}_n^2}{\sigma^2}\right) = 2(n-1)$

$$\frac{(n-1)^2}{\sigma^4}\text{Var}(\hat{\sigma}_n^2) = 2(n-1)$$

$$\text{Var}(\hat{\sigma}_n^2) = \frac{2\sigma^4}{n-1}$$

And $E\left(\dfrac{(n-1)\hat{\sigma}_n^2}{\sigma^2}\right) = n-1$

$$\frac{(n-1)}{\sigma^2}E(\hat{\sigma}_n^2) = n-1$$

$$E(\hat{\sigma}_n^2) = 0$$

So $\hat{\sigma}_n^2 \sim N\left(0, \dfrac{2\sigma^4}{n-1}\right)$

Then $\hat{\sigma}_n^2 - \sigma^2 \sim N\left(0, \dfrac{2\sigma^4}{n-1}\right)$

$$\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2) \sim N\left(0, \frac{n2\sigma^4}{n-1}\right)$$

Since n is large, we approximate $\dfrac{n}{n-1} \approx 1$

Hence $\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2) \sim N(0, 2\sigma^4)$

Then, using delta method,

$$\sqrt{n}\,(g(\hat{\sigma}_n^2) - g(\sigma^2)) \sim N(0, g'(\sigma^2)^2 \cdot 2\sigma^4)$$

Let $g(x) = \psi(\sqrt{x}) = \Phi\left(\frac{\sqrt{x}}{2}\right)$

Then $g(\sigma^2) = \psi(\sigma) = \Phi\left(\frac{\sigma}{2}\right)$

Also, $\frac{d}{d\sigma} g(\sigma^2) = g'(\sigma^2)$

$= \left(\Phi'\left(\frac{\sigma}{2}\right) \cdot \frac{1}{4\sigma}\right)$

$= \frac{1}{4\sigma} \times$ PDF of $N(0,1)$ at $\frac{\sigma}{2}$

$= \frac{1}{4\sigma} \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\sigma}{2}\right)^2\right)$

Then, $g'(\sigma^2)^2 \cdot 2\sigma^4 = \left[\frac{1}{4\sigma} \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\sigma}{2}\right)^2\right)\right]^2 \cdot 2\sigma^4$

$= \frac{1}{16\sigma^2} \times \frac{1}{2\pi} \exp\left(-\frac{\sigma^2}{4}\right) \times 2\sigma^4$

$= \frac{\sigma^2}{16\pi} \exp\left(-\frac{\sigma^2}{4}\right)$

$= 0.01207746$

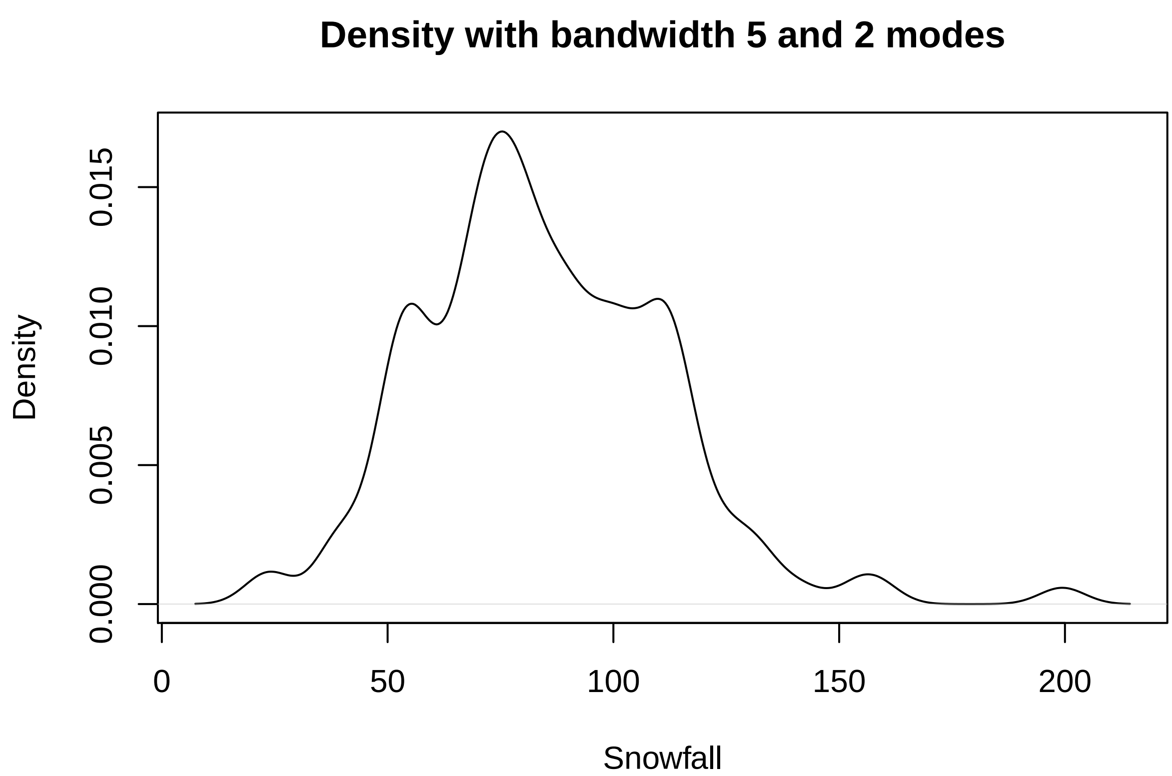
STA355 A2

Harold Hyun Woo Lee

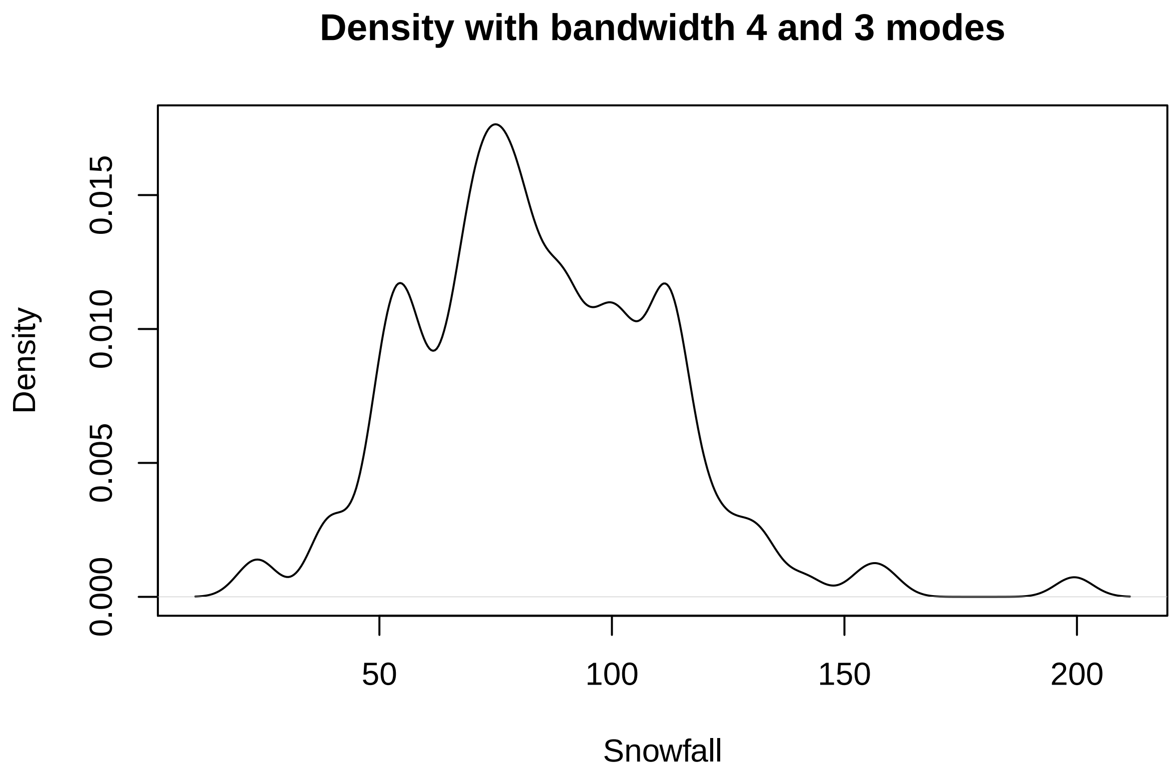
2/21/2021

Q1a)

```
dat <- scan("buffsnow.txt")

plot(density(dat, bw=5), main="Density with bandwidth 5 and 2 modes", xlab="Snowfall")
```




```
plot(density(dat, bw=4), main="Density with bandwidth 4 and 3 modes", xlab="Snowfall")
```


Density with bandwidth 4 and 3 modes
Density
0.000
0.005
0.010
0.015
50
100
150
200
Snowfall

Q1c)

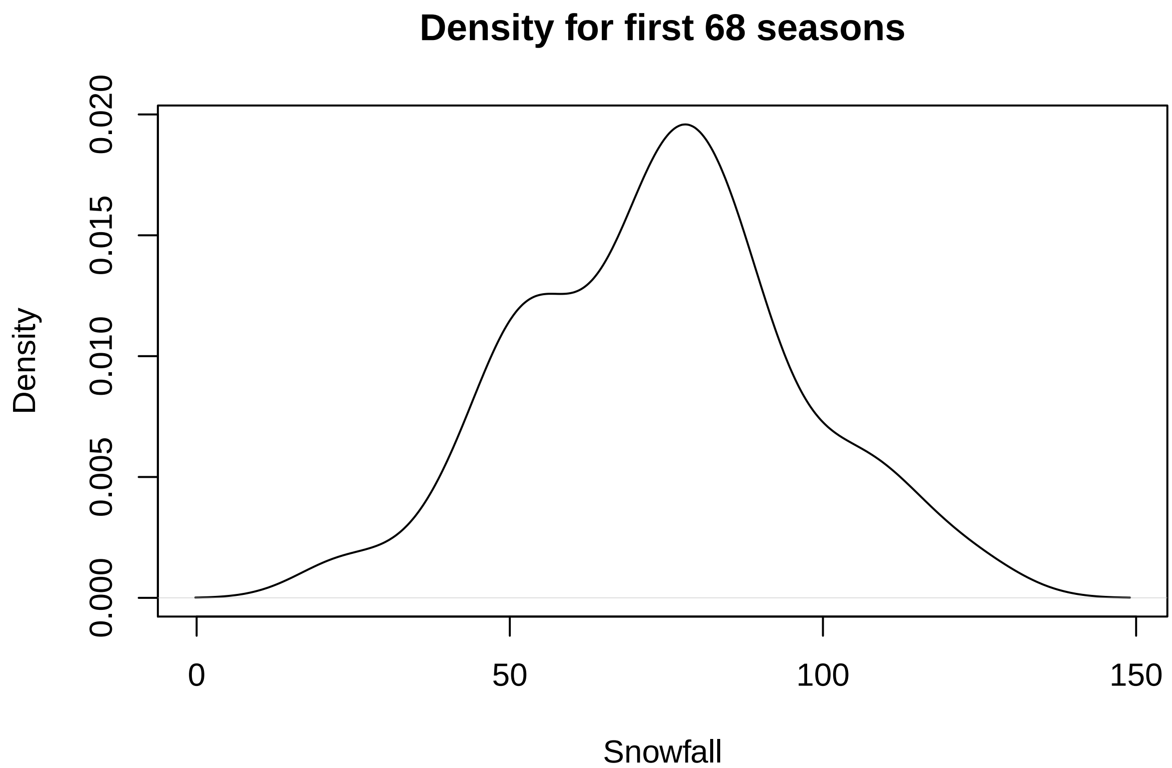
```r
kde.cv <- function(x,h) {
             n <- length(x)
             if (missing(h)) {
                r <- density(x)
                h <- r$bw/4 + 3.75*c(0:100)*r$bw/100
                }
             cv <- NULL
             for (j in h) {
                cvj <- 0
                for (i in 1:n) {
                   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
              }

# Split the data into two
dat1 <- dat[1:68]
dat2 <- dat[69:136]

# Compute bandwidth for dat1
r1 <- kde.cv(dat1)
#plot(r1$bw,r1$cv) # plot of bandwidth versus CV
r1$bw[r1$cv==max(r1$cv)] # bandwidth maximizing CV
```

```
## [1] 7.546673
```

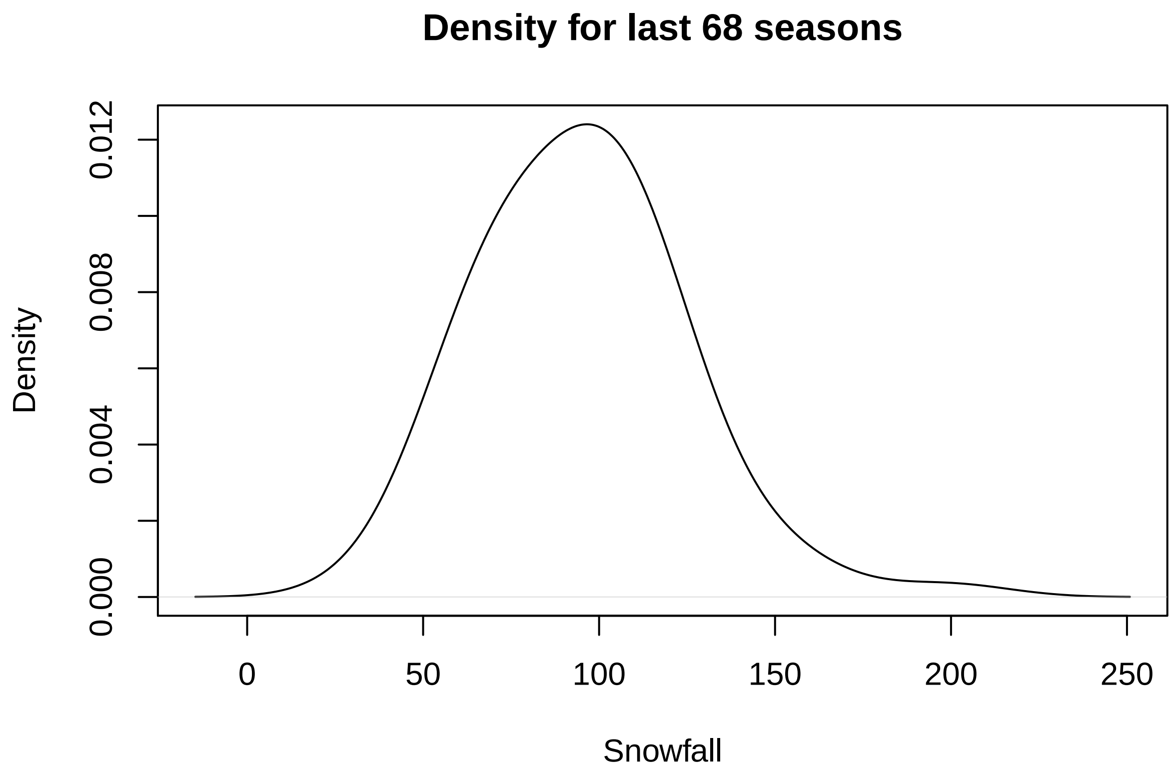
```r
# Compute bandwidth for dat2
r2 <- kde.cv(dat2)
#plot(r2$bw,r2$cv) # plot of bandwidth versus CV
r2$bw[r2$cv==max(r2$cv)] # bandwidth maximizing CV
```

```
## [1] 17.15592
```

```r
plot(density(dat1, bw=7.54), main="Density for first 68 seasons", xlab="Snowfall")
```



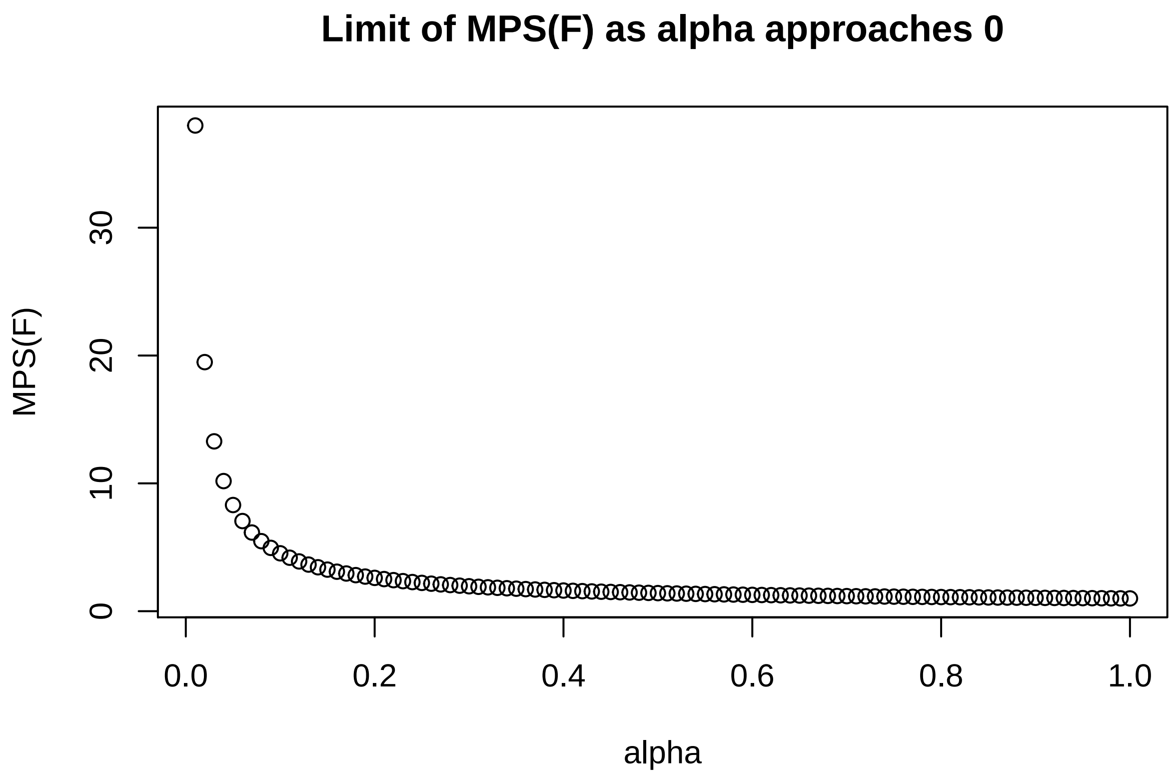

```
plot(density(dat2, bw=17.15), main="Density for last 68 seasons", xlab="Snowfall")
```



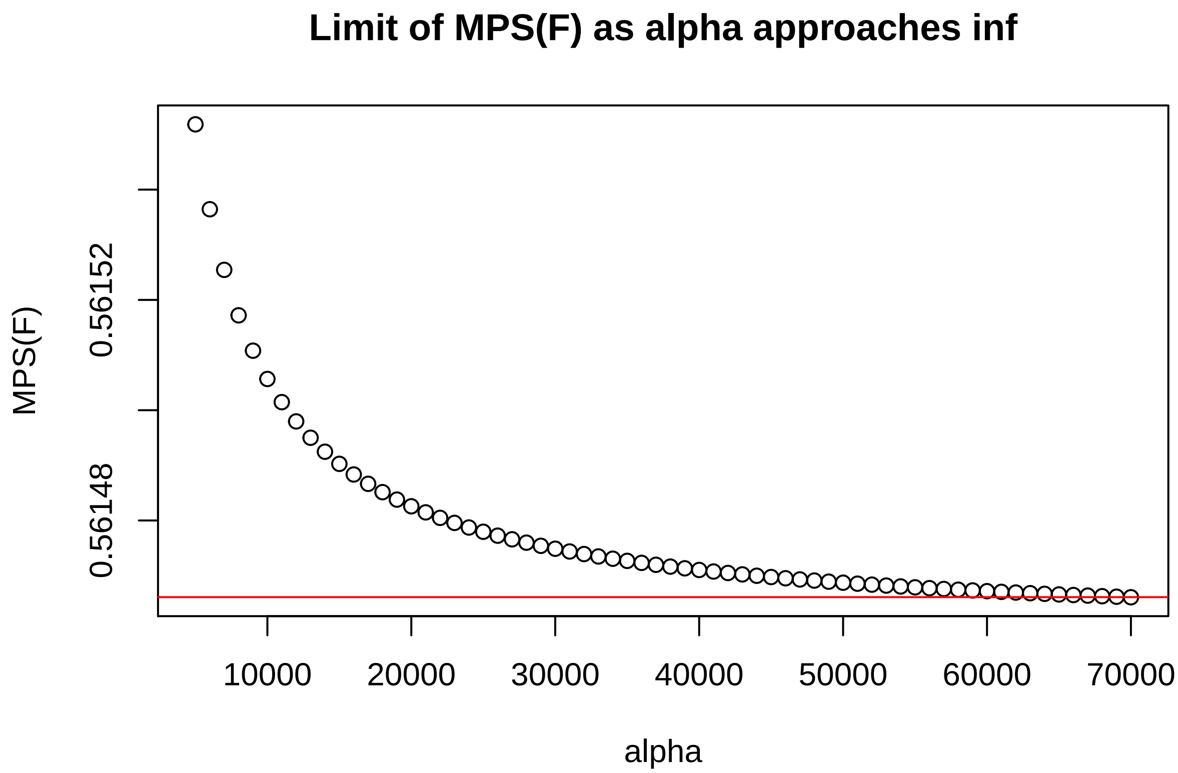
The estimated density for the first 68 seasons is approximately 0.02, and the estimated density for the last 68 seasons is approximately 0.012. We see that the bandwidths are 7.54 and 17.15 respectively. The later half of the data is somewhat more dispersed and has a larger bandwidth value. In addition, we see that the density for the first 68 seasons is much higher than the density for the later half.

Q2b)

```r
x1 <- seq(0.01, 1, by=0.01)
plot(x1, gamma(1/x1 + 1)^x1, main="Limit of MPS(F) as alpha approaches 0",
     xlab = "alpha", ylab="MPS(F)")
```




```r
x2 <- seq(5000, 70000, by = 1000)
plot(x2, gamma(1/x2 + 1)^x2, main="Limit of MPS(F) as alpha approaches inf",
     xlab = "alpha", ylab="MPS(F)")
abline(h = 0.5614661, col="red")
```




```r
tail(gamma(1/x2 + 1)^x2, 1)
```

```
## [1] 0.5614661
```

The 2 graphs above shows that as $\alpha \to 0$, $MPS(F) \to \inf$.
In addition, as $\alpha \to \inf$, $MPS(F) \to 0.5614$.

Q2c)

```r
income <- scan("incomes.txt")


x_bar <- mean(income)
count = c()
for (val in income){
  if (val < x_bar){
    count = c(count, 1)
  }
  if (val > x_bar){
    count = c(count, 0)
  }
}
MPS_hat = sum(count)/200


loo <- NULL
for (i in 1:200){
  loo <- c(loo, mean(count[-i]))
}
sehat <- sqrt(199*sum((loo-mean(loo))^2)/200)
sehat
```

```
## [1] 0.02631229
```

```r
MPS_hat
```

```
## [1] 0.835
```

Here we see that $M\hat{PS}(F) = 0.835$ and the estimate of the standard error using jackknife is 0.026

2d)

```
sigma_2 = var(log(income))

var_MPS = sigma_2/(16*pi) * exp(-sigma_2/4)

SE = sqrt(var_MPS)/sqrt(200)
SE
```

```
## [1] 0.01207746
```

The estimate of the standard error of $\psi(\hat{\sigma})$ for the income data is about 0.0120. This is smaller than the estimate found in part C.