Solutions: Assignment #2 STA355H1S

1 (a) Plots of density estimates for bandwidths 0.0039 (the default for `density`), 0.008, 0.002, and 0.001 are shown on the following page with black, blue, red, and cyan lines respectively. Note that decreasing the bandwidth makes local modes somewhat more evident. The R code for the plots is given below:

```
> plot(density(stamp),ylim=c(0,70),lwd=4,main=" ")
> lines(density(stamp,bw=0.008),col="blue",lwd=4)
> lines(density(stamp,bw=0.002),col="red",lwd=4)
> lines(density(stamp,bw=0.001), col="cyan",lwd=4)
> points(stamp,rep(0,486),pch=20,cex=0.5) # draw points on x-axis
```

Counting modes is a tricky business and what constitutes a mode is very subjective. However, one can argue that there are certainly 5 modes for $h = 0.002$ and possibly 7 modes at $h = 0.001$.

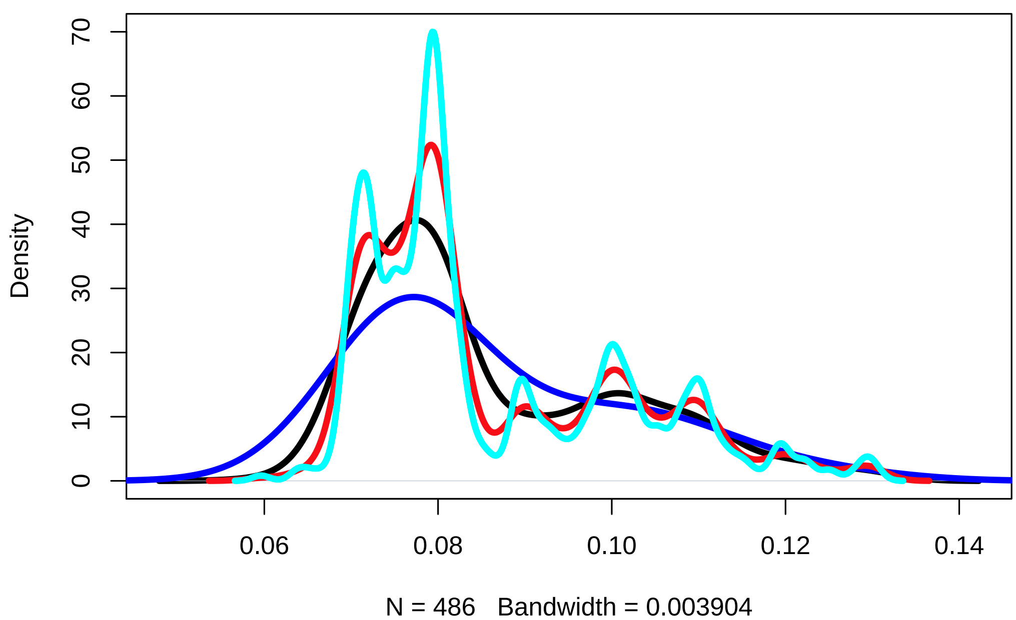

Figure 1: Density estimates from part (a).

(b) Note that

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

$$= \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{1}{nh}\sum_{j\neq i} w\left(\frac{X_i - X_j}{h}\right) + \frac{1}{nh}w(0)\right)$$

Using the conditions $w(0) > 0$ and $h^{-1}w(x/h) \to 0$ as $h \downarrow 0$ for $x \neq 0$, it follows that

$$\frac{1}{nh}\sum_{j\neq i} w\left(\frac{X_i - X_j}{h}\right) \to 0$$

and

$$\frac{1}{nh}w(0) \to \infty$$

as $h \downarrow 0$. Thus

$$\frac{1}{nh}\sum_{j\neq i} w\left(\frac{X_i - X_j}{h}\right) + \frac{1}{nh}w(0) \to \infty$$

as $h \downarrow 0$ for each i and so $\mathcal{L}(h) \to \infty$.
Likewise for $\mathrm{CV}(h)$, we have

$$\frac{1}{(n-1)h}\sum_{j\neq i} w\left(\frac{X_i - X_j}{h}\right) \to 0$$

as $h \downarrow 0$ for each i (since $X_i - X_j \neq 0$ for $i \neq j$ and so $\mathrm{CV}(h) \to -\infty$ as $h \downarrow 0$. As $h \uparrow \infty$,

$$\frac{1}{(n-1)h} w\left(\frac{X_i - X_j}{h}\right) \to 0$$

and so $\mathrm{CV}(h) \to -\infty$ as $h \uparrow \infty$.

(c) Figure 2 shows a plot of $\mathrm{CV}(h)$ versus h with $\mathrm{CV}(h)$ maximized at $h = 0.00115$; the resulting density estimate is shown in Figure 3. As with the estimates in Figure 1, the number of modes is quite subjective — there appear to be 5 obvious modes with (possibly) another 2 smaller modes for larger values. The R code is given below:

```
> r <-  kde.cv(stamp)
> plot(r$bw,r$cv,pch=20)
> lines(r$bw,r$cv,lwd=0.5)
> r$bw[r$cv==max(r$cv)]
[1] 0.001151797
> abline(v=0.001151797,lwd=2,col="red")
> plot(density(stamp,bw=0.001151797),lwd=4,main=" ")
> points(stamp,rep(0,486),pch=20,cex=0.5) # draw points on x-axis
```

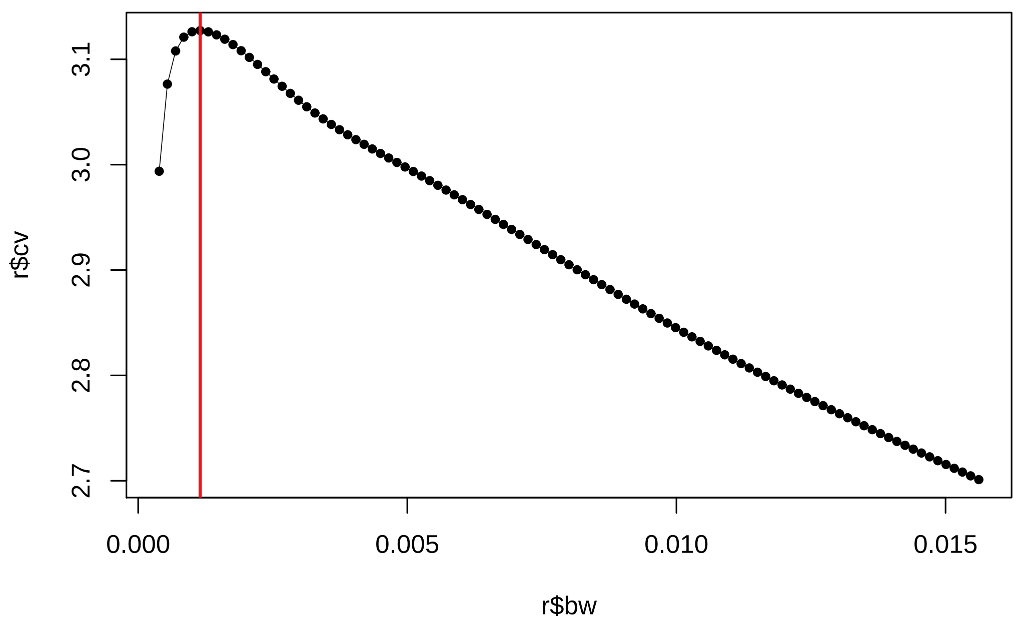

Figure 2: Plot of CV(h) versus h; the maximum occurs at $h = 0.00115$.

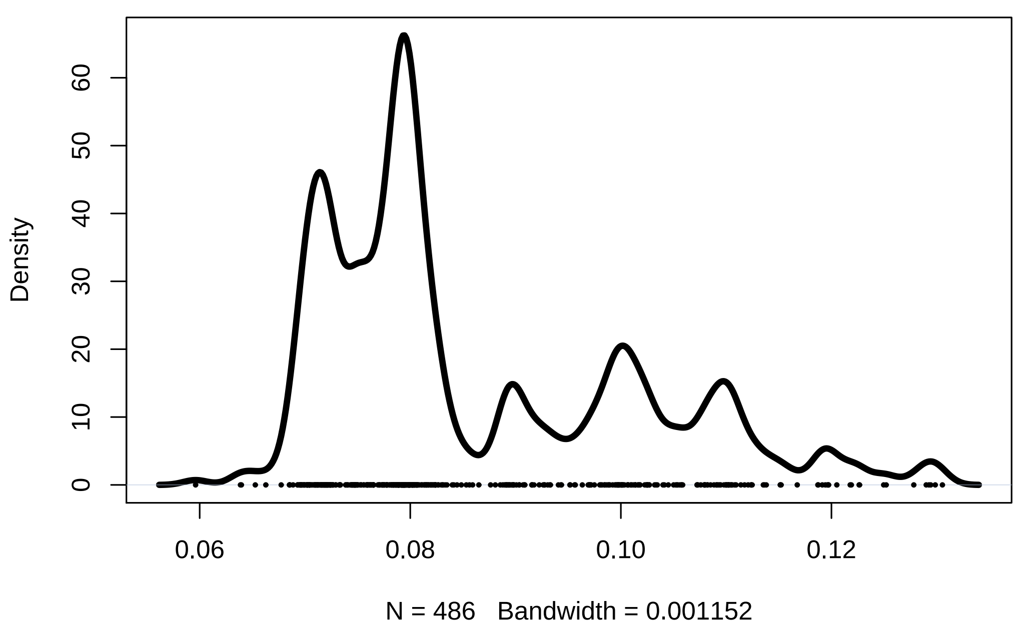

Figure 3: Density estimate using the bandwidth maximizing CV(h).

2. (a) Since $\mathcal{L}_F(t) = \dfrac{1}{\mu(F)} \displaystyle\int_0^t F^{-1}(s)\, ds$, we have

$$\mathcal{L}'_F(t) = \frac{1}{\mu(F)} F^{-1}(t)$$

Thus $\mathcal{L}'_F(t) = 1$ implies that $F^{-1}(t) = \mu(F)$ or $t = F(\mu(F)) = \mathrm{MPS}(F)$.

(b) $\mathcal{L}'_F(t) = (\alpha + 1)t^\alpha$ and setting $\mathcal{L}'_F(\mathrm{MPS}(F)) = 1$, we get

$$\mathrm{MPS}(F) = (1 + \alpha)^{-1/\alpha}.$$

Since $\mathrm{MIS}(F) = \mathcal{L}_F(\mathrm{MPS}(F))$, we get

$$\mathrm{MIS}(F) = (1 + \alpha)^{-(1+\alpha)/\alpha}.$$

(c) The following simple function computes MPS:

```
MPS <- function(x) {
         sum(x<mean(x))/length(x)
         }
```

For the incomes data, we get

```
> x <- scan("incomes.txt")
> > MPS(x)
[1] 0.69
```

The jackknife standard error can be evaluated as follows:

```
> loo <- NULL
> for (i in 1:200) {
+    xi <- x[-i]
+    loo <- c(loo,MPS(xi))
+    }
> se <- sqrt(199*sum((loo-mean(loo))^2)/200)
> se
[1] 0.07811778
```

(d) The following R function estimates $\mathcal{L}_F(t)$ for a given value of t:

```
lorenz <- function(x,t) {
            n <- length(x)
            x <- sort(x)
            m <- ceiling(n*t)
            sum(x[1:m])/sum(x)
            }
```

We can then use this function with the function `MPS` defined earlier to estimate $\text{MIS}(F)$:

```
> MIS <- lorenz(x,MPS(x))
> MIS
[1] 0.3480396
```

We can now compute the jackknife standard error as follows:

```
> loo <- NULL
> for (i in 1:200) {
+    xi <- x[-i]
+    loo <- c(loo,lorenz(xi,MPS(xi)))
+    }
> se <- sqrt(199*sum((loo-mean(loo))^2)/200)
> se
[1] 0.08170359
```

Note that we need to input the leave-one-out MPS estimate `MPS(xi)` in order to compute the leave-one-out MIS estimate.

Aside: It is possible to derive the approximate distribution of the estimator

$$\widehat{\text{MPS}}(F) = \frac{1}{n}\sum_{i=1}^{n} I(X_i < \bar{X})$$

assuming that F is a continuous cdf with pdf $f = F'$. The idea is the following:

$$\begin{aligned}\sqrt{n}(\widehat{\text{MPS}}(F) - \text{MPS}(F)) &= \sqrt{n}(\widehat{\text{MPS}}(F) - F(\mu)) \\ &= \sqrt{n}(\widehat{\text{MPS}}(F) - F(\bar{X})) + \underbrace{\sqrt{n}(F(\bar{X}) - F(\mu))}_{\approx f(\mu)\sqrt{n}(\bar{X}-\mu)}.\end{aligned}$$

It is also possible to show that

$$\sqrt{n}(\widehat{\text{MPS}}(F) - F(\bar{X})) \approx \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i < \mu) - F(\mu)\right)$$

in the sense that the difference between the two tends in probability to 0 as $n \to \infty$. From this, we have

$$\sqrt{n}(\widehat{\text{MPS}}(F) - \text{MPS}(F)) \approx \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i < \mu) - F(\mu)\right) + f(\mu)\sqrt{n}(\bar{X} - \mu),$$

which suggests that

$$\text{Var}(\widehat{\text{MPS}}(F)) \approx \frac{1}{n}\left\{F(\mu)(1 - F(\mu)) + f^2(\mu)\text{Var}(X_i) + 2f(\mu)\text{Cov}(I(X_i < \mu), X_i)\right\}.$$

It is tempting to estimate the standard error of $\widehat{\text{MPS}}(F)$ as

$$\widehat{\text{se}}(\widehat{\text{MPS}}(F)) = \left(\frac{\widehat{\text{MPS}}(F)(1 - \widehat{\text{MPS}}(F))}{n}\right)$$

(which would be correct if we knew μ) but we do need to account for the uncertainty in the estimation of $\bar{X}$ in the standard error.

Supplemental problems:

3. (a) For the "leave-one-out" samples, we are still trimming r observations at each end and averaging the remaining $n - 2r - 1$ observations. If one of $X_{(1)}, \cdots, X_{(r)}$ is deleted then $X_{(r+1)}$ becomes the r smallest in the "leave-one-out" sample and so

$$\widehat{\theta}_{-i}(\alpha) = \frac{1}{n - 2r - 1}\sum_{j=r+2}^{n-r} X_{(j)} \quad \text{for } i = 1, \cdots, r$$

Likewise, if one of $X_{(n-r+1)}, \cdots, X_{(n)}$ is deleted that $X_{(n-r)}$ becomes the r largest in the "leave-one-out" sample and so

$$\widehat{\theta}_{-i}(\alpha) = \frac{1}{n - 2r - 1}\sum_{j=r+1}^{n-r-1} X_{(j)} \quad \text{for } i = n - r + 1, \cdots, n$$

Finally, if $r + 1 \le i \le n - r$ and $X_{(i)}$ is deleted then the r largest and smallest remain the same and so

$$\widehat{\theta}_{-i}(\alpha) = \frac{1}{n - 2r - 1}\sum_{\substack{j=r+1 \\ j \neq i}}^{n-r} X_{(j)} \quad \text{for } i = r + 1, \cdots, n - r$$

(b) For each i, the pseudo-values are defined by

$$\begin{aligned}\Phi_i &= n\widehat{\theta}(\alpha) - (n-1)\widehat{\theta}_{-i}(\alpha) \\ &= \frac{n}{n - 2r}\sum_{j=r+1}^{n-r} X_{(j)} - \frac{n-1}{n - 2r - 1}\sum_{\substack{j=r+1 \\ j \neq \ell(i)}}^{n-r} X_{(j)}\end{aligned}$$

where

$$\ell(i) = \begin{cases} r+1 & \text{for } i = 1, \cdots, r+1 \\ i & \text{for } i = r+2, \cdots, n-r-1 \\ n-r & \text{for } i = n-r, \cdots, n. \end{cases}$$

Now note that

$$\frac{n-1}{n-2r-1} \sum_{\substack{j=r+1 \\ j \neq \ell(i)}}^{n-r} X_{(j)} = \frac{n-1}{n-2r-1} \left\{ \sum_{j=r+1}^{n-r} X_{(j)} - X_{(\ell(i))} \right\}.$$

Thus

$$\Phi_i = \left(\frac{n}{n-2r} - \frac{n-1}{n-2r-1} \right) \sum_{j=r+1}^{n-r} X_{(j)} + \frac{n-1}{n-2r-1} X_{(\ell(i))}$$

or

$$\begin{aligned} \Phi_i &= \frac{n-1}{n-2r-1} X_{(r+1)} - \frac{2r}{(n-2r)(n-2r-1)} \sum_{k=r+1}^{n-r} X_{(k)} \quad \text{for } i = 1, \cdots, r+1 \\ \Phi_i &= \frac{n-1}{n-2r-1} X_{(i)} - \frac{2r}{(n-2r)(n-2r-1)} \sum_{k=r+1}^{n-r} X_{(k)} \quad \text{for } i = r+2, \cdots, n-r-1 \\ \Phi_i &= \frac{n-1}{n-2r-1} X_{(n-r)} - \frac{2r}{(n-2r)(n-2r-1)} \sum_{k=r+1}^{n-r} X_{(k)} \quad \text{for } i = n-r, \cdots, n. \end{aligned}$$

The jackknife variance estimator is simply

$$\widehat{\mathrm{Var}}(\widehat{\theta}(\alpha)) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\Phi_i - \bar{\Phi})^2$$

Since $\Phi_i = (n-1)X_{(\ell(i))}/(n-2r-1) + \text{constant}$, we can write

$$\widehat{\mathrm{Var}}(\widehat{\theta}(\alpha)) = \frac{n-1}{n(n-2r-1)^2} \sum_{i=1}^{n} (X_{(\ell(i))} - \widehat{\omega})^2$$

where

$$\widehat{\omega} = \frac{1}{n} X_{(\ell(i))} = \frac{1}{n} \left\{ (r+1) X_{(r+1)} + \sum_{i=r+2}^{n-r-1} X_{(i)} + (r+1) X_{(n-r)} \right\}.$$

($\widehat{\omega}$ is called a **Winsorized mean**.)

Aside: Assuming the jackknife variance estimator is reasonable for each r, we could use it to choose a "good" value of r by minimizing the variance; the following function computes estimates for a range of values for r:

```
winsor <- function(x,r) {
            x <- sort(x)
            n <- length(x)
            if (missing(r)) r <- floor(0.4*n)
            vars <- var(x)/n # jackknife variance for sample mean (r=0)
```

```
            for (i in 1:r) {
               # Winsorized sample
               xw <- c(rep(x[i+1],i),x[(i+1):(n-i)],rep(x[n-i],i))
               wm <- mean(xw) # Winsorized mean
               # jackknife variance estimate
               vars <- c(vars,(n-1)*sum((xw-wm)^2)/(n*(n-2*i-1)^2))
               }
            list(var=vars,r=c(0:r))
            }
```

Plots for Normal, Logistic, and Student's t with 3 degrees of freedom for $n = 100$ (using the code given below) are given in Figure 4. The plot for the Normal distribution seems to clearly suggest that no trimming (or at most modest trimming) is best; the plots for the other two distributions are less conclusive. However, these plots do seem to give more definitive evidence about how to choose r when n is larger.

```
> x1 <- rnorm(100)
> x2 <- rlogis(100)
> x3 <- rt(100,3)
> r1 <- winsor(x1,45)
> r2 <- winsor(x2,45)
> r3 <- winsor(x3,45)
> plot(r1$r,r1$var,xlab="r",ylab="jackknife variance",main="Normal")
> plot(r2$r,r2$var,xlab="r",ylab="jackknife variance",main="Logistic")
> plot(r3$r,r3$var,xlab="r",ylab="jackknife variance",
+ main="Student's t with 3 df")
```

4. (a) $E_\theta(\widetilde{\theta}) = aE_\theta(\widehat{\theta}_1) + (1-a)E_\theta(\widehat{\theta}_2) = a\,\theta + (1-a)\,\theta = \theta$.

(b) We want to find a to minimize

$$\mathrm{Var}(\widetilde{\theta}) = a^2\mathrm{Var}(\widehat{\theta}_1) + (1-a)^2\mathrm{Var}(\widehat{\theta}_2) + 2a(1-a)\mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2)$$

Differentiating and solving for a, we get

$$a = \frac{\mathrm{Var}(\widehat{\theta}_2) - \mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2)}{\mathrm{Var}(\widehat{\theta}_1) + \mathrm{Var}(\widehat{\theta}_2) - 2\mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2)}.$$

$a = 1$ if $\mathrm{Var}(\widehat{\theta}_1) = \mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2)$.
Note that $a < 0$ if $\mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2) > \mathrm{Var}(\widehat{\theta}_2)$. Note that this condition implies that $\mathrm{Var}(\widehat{\theta}_1) > \mathrm{Var}(\widehat{\theta}_2)$ and typically we need $\mathrm{Var}(\widehat{\theta}_1)$ to be much bigger than $\mathrm{Var}(\widehat{\theta}_2)$ in order for $\mathrm{Cov}(\widehat{\theta}_1, \widehat{\theta}_2) > \mathrm{Var}(\widehat{\theta}_2)$.

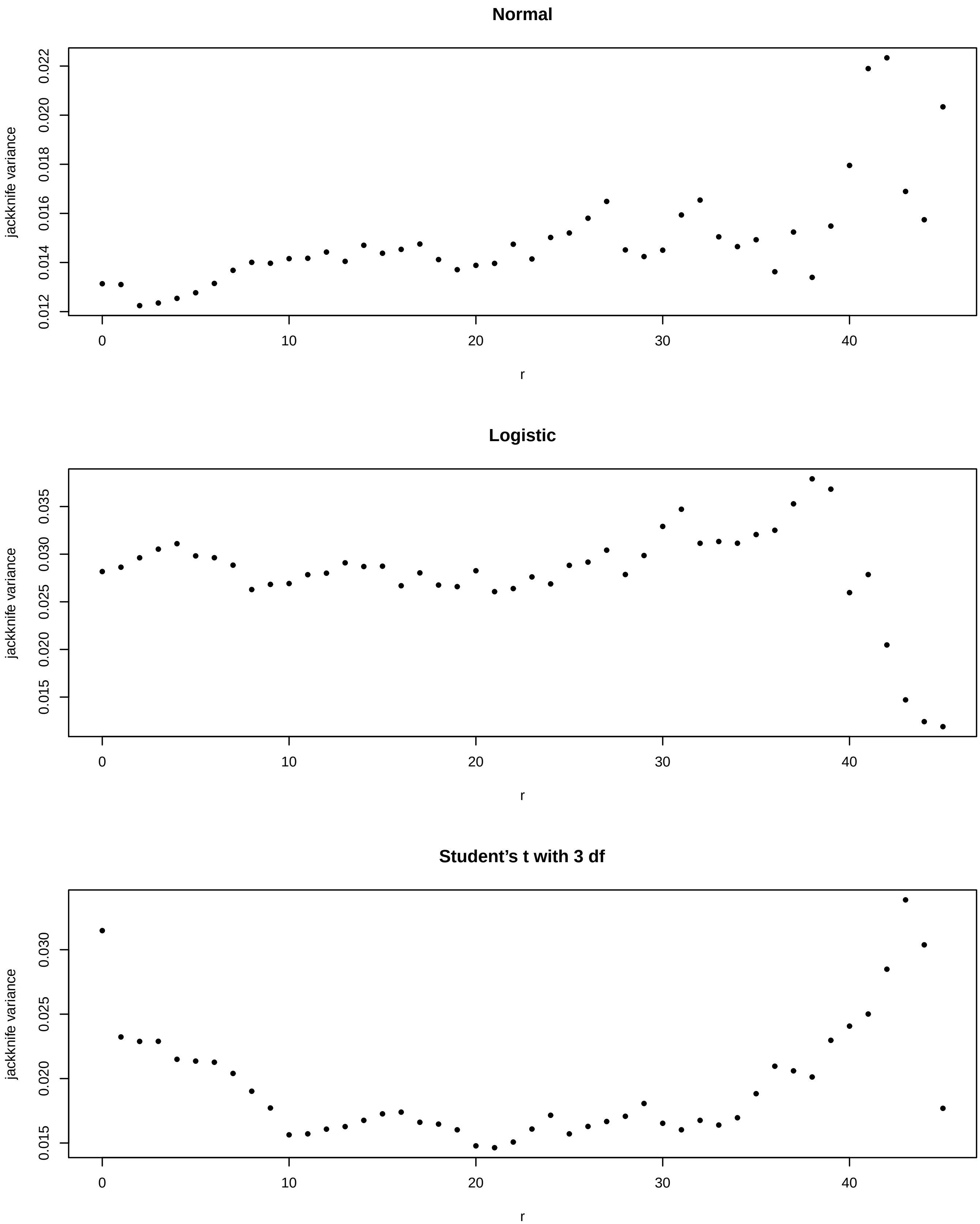

Figure 4: Plots of jackknife variances as a function of r for the Normal, Logistic, and Student's t (3 df) distributions for $n = 100$.

5. (a) We must show that (i) $\widehat{f}(x) \geq 0$ for all x, and (ii) $\int_{-\infty}^{\infty} \widehat{f}(x)\, dx = 1$; clearly $\widehat{f}(x) \geq 0$. To show (ii), note that

$$\begin{aligned}
\int_{-\infty}^{\infty} \widehat{f}(x)\, dx &= \sum_{j=1}^{k} \int_{a_{j-1}}^{a_j} \widehat{f}(x)\, dx \\
&= \sum_{j=1}^{k} \left\{ (a_j - a_{j-1}) \times \frac{1}{n(a_j - a_{j-1})} \sum_{i=1}^{n} I(a_{j-1} \leq X_i < a_j) \right\} \\
&= \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n} I(a_{j-1} \leq X_i < a_j) \\
&= 1.
\end{aligned}$$

(b) Suppose that $a_{j-1} \leq x < a_j$. Then

$$\begin{aligned}
E[\widehat{f}(x)] &= \frac{1}{n(a_j - a_{j-1})} \sum_{i=1}^{n} P(a_{j-1} \leq X_i < a_j) \\
&= \frac{F(a_j) - F(a_{j-1})}{a_j - a_{j-1}}.
\end{aligned}$$

Thus we can express the bias as

$$E[\widehat{f}(x)] - f(x) = \int_{a_{j-1}}^{a_j} \frac{f(t) - f(x)}{a_j - a_{j-1}}\, dt.$$

Likewise,

$$\begin{aligned}
\mathrm{Var}[\widehat{f}(x)] &= \frac{1}{n^2(a_j - a_{j-1})^2} \sum_{i=1}^{n} \mathrm{Var}[I(a_{j-1} \leq X_i < a_j)] \\
&= \frac{\{F(a_j) - F(a_{j-1})\}\{1 - F(a_j) + F(a_{j-1}\}}{n(a_j - a_{j-1})^2}
\end{aligned}$$

(c) Assuming the density f is continuous and positive at x, the bias of $\widehat{f}(x)$ for $x \in [a_{j-1}, a_j)$ will tend to 0 if $a_j - a_{j-1} \to 0$. On the other hand, if $a_j - a_{j-1}$ is small and $x \in [a_{j-1}, a_j)$ then

$$\mathrm{Var}[\widehat{f}(x)] \approx \frac{f(x)}{n(a_j - a_{j-1})}$$

since

$$\frac{F(a_j) - F(a_{j-1})}{a_j - a_{j-1}} \approx f(x).$$

Thus $\mathrm{Var}[\widehat{f}(x)] \to 0$ if $n(a_j - a_{j-1}) \to \infty$.
Therefore to have both the bias and variance tending to 0 as $n \to \infty$, we require $a_j - a_{j-1} \to 0$ and $n(a_j - a_{j-1}) \to \infty$.

6. (a) $g'(t) = 1 - F^{-1}(t)/\mu(F)$ and note that $g'(t)$ is a non-increasing function of t since $F^{-1}(t)$ is a non-decreasing function. Therefore $g(t)$ is maximized at t satisfying $g'(t) = 0$ or $F^{-1}(t) = \mu(F)$. If $F^{-1}(t)$ is strictly increasing for $t \in (0, 1)$, g is maximized at $t = F(\mu(F))$.

(b) From part (a),

$$\mathcal{P}(F) = F(\mu(F)) - \frac{1}{\mu(F)} \int_0^{F(\mu(F))} .F^{-1}(s)\, ds$$

Making the change of variable $x = F^{-1}(s)$ so that $s = F(x)$ and $ds = f(x)\, dx$, we get

$$\begin{aligned}
\mathcal{P}(F) &= F(\mu(F)) - \frac{1}{\mu(F)} \int_0^{F(\mu(F))} .F^{-1}(s)\, ds \\
&= \int_0^{\mu(F)} f(x)\, dx - \frac{1}{\mu(F)} \int_0^{\mu(F)} x f(x)\, dx \\
&= \frac{1}{\mu(F)} \int_0^{\mu(F)} (\mu(F) - x) f(x)\, dx \\
&= \frac{1}{2\mu(F)} \int_0^{\infty} |x - \mu(F)| f(x)\, dx
\end{aligned}$$

since

$$\begin{aligned}
\int_0^{\infty} |x - \mu(F)| f(x)\, dx &= \int_0^{\mu(F)} (\mu(F) - x) f(x)\, dx + \int_{\mu(F)}^{\infty} (x - \mu(F)) f(x)\, dx \\
\text{and} \quad \int_0^{\mu(F)} (\mu(F) - x) f(x)\, dx &= \int_{\mu(F)}^{\infty} (x - \mu(F)) f(x)\, dx.
\end{aligned}$$

(c) A substitution principle estimator for $\mathcal{P}(F)$ is

$$\widehat{\mathcal{P}} = \frac{1}{2\bar{X}} \left\{ \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}| \right\}$$

using $\bar{X}$ as a substitution principle estimator of $\mu(F)$. The leave-one-out estimators are then

$$\widehat{\mathcal{P}}_{-i} = \frac{1}{2\bar{X}_{-i}} \left\{ \frac{1}{n-1} \sum_{j \neq i} |X_j - \bar{X}_{-i}| \right\}$$

where

$$\bar{X}_{-i} = \frac{1}{n-1} \sum_{j \neq i} X_j.$$

The following R function computes the estimate of the Pietra index and its jackknife standard error:

```
pietra <- function(x,jackknife=T) {
            n <- length(x)
            est <- mean(abs(x-mean(x)))/(2*mean(x))
```

```
        if (jackknife) {
           loo <- NULL
           for (i in 1:n) {
             xi <- x[-i]
             loo <- c(loo, mean(abs(xi-mean(xi)))/(2*mean(xi)))
             }
           se <- sqrt(sum((loo-mean(loo))^2)*(n-1)/n)
           r <- list(index=est,std.err=se)
           }
        else {
           r <- list(index=est)
           }
        r
        }
```

For the incomes data, we get

```
> inc <- scan("incomes.txt")
> r <- pietra(inc)
> r
$index
[1] 0.3419604
$std.err
[1] 0.02150457
```

The estimate is 0.342 and its estimated standard error is 0.022.

(d) I believe there is a typo in the problem, If $\ln(X_i) \sim \mathcal{N}(\mu, \sigma^2)$ then

$$E(X_i) = E[\exp(\ln(X_i))] = \exp(\mu)\exp(\sigma^2/2) = \mu(F).$$

Then $Y_i = X_i/\mu(F)$ has a log-normal distribution with parameters $-\sigma^2/2$ and σ^2 so that $E(Y_i) = E[X_i/\mu(F)] = 1$. Thus $\mathcal{P}(F) = E(|Y_i - 1|)/2$. This can be evaluated as follows (skipping some steps!):

$$\begin{aligned}
\mathcal{P}(F) &= \frac{1}{2}\left\{\int_{-\infty}^{0}(1-\exp(y))\frac{1}{\sigma}\phi\left(\frac{(y+\sigma^2/2)^2}{2\sigma^2}\right)dy + \int_0^{\infty}(\exp(y)-1)\frac{1}{\sigma}\phi\left(\frac{(y+\sigma^2/2)^2}{2\sigma^2}\right)dy\right\} \\
&= \int_0^{\infty}(\exp(y)-1)\frac{1}{\sigma}\phi\left(\frac{(y+\sigma^2/2)^2}{2\sigma^2}\right)dy \\
&= \int_0^{\infty}\frac{1}{\sigma}\phi\left(\frac{(y-\sigma^2/2)^2}{2\sigma^2}\right)dy - \int_0^{\infty}\frac{1}{\sigma}\phi\left(\frac{(y+\sigma^2/2)^2}{2\sigma^2}\right)dy \\
&= 1-\Phi(-\sigma/2) - [1-\Phi(\sigma/2)] \\
&= 2\Phi(\sigma/2) - 1.
\end{aligned}$$