Homework 7

July 19, 2021

Problem 1

(1)

Need to use the Newton-Raphson algorithm to try and minimize:

$$l(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - exp(\theta x_i))^2$$

So I used the chain rule with $f(\theta) = \frac{1}{2}\sum_{i=1}^{n}(x)^2$ and $g(\theta) = y_i - e^{\theta x_i}$

I found $f'(\theta) = \sum_{i=1}^{n}(x)$ and $g'(\theta) = -x_i e^{\theta x_i}$

And got:

$$l'(\theta) = -\sum_{i=1}^{n} x_i(y_i - e^{\theta x_i})e^{\theta x_i}$$

For the second derivative I used the product rule with $f(\theta) = y_i - e^{\theta x_i}$ and $g(\theta) = e^{\theta x_i}$

For which the derivatives are: $f'(\theta) = -x_i e^{\theta x_i}$ and $g'(\theta) = x_i e^{\theta x_i}$

And got:

$$\begin{aligned}
l''(\theta) &= -\sum_{i=1}^{n} x_i[-x_i e^{\theta x_i} e^{\theta x_i} + (y_i - e^{\theta x_i})x_i e^{\theta x_i}] \\
&= -\sum_{i=1}^{n} x_i^2[-e^{\theta x_i} e^{\theta x_i} + (y_i - e^{\theta x_i})e^{\theta x_i}] \\
&= -\sum_{i=1}^{n} x_i^2 e^{\theta x_i}[-e^{\theta x_i} + (y_i - e^{\theta x_i})] \\
&= -\sum_{i=1}^{n} x_i^2 e^{\theta x_i}[(y_i - 2e^{\theta x_i})] \\
&= \sum_{i=1}^{n} x_i^2 e^{\theta x_i}(2e^{\theta x_i} - y_i)
\end{aligned}$$

So the algorithm ends up looking like:

$$\theta(t+1) = \theta(t) - \frac{-\sum_{i=1}^{n} x_i(y_i - e^{\theta x_i})e^{\theta x_i}}{\sum_{i=1}^{n} x_i^2 e^{\theta x_i}(2e^{\theta x_i} - y_i)}$$

I coded this in R:

```r
#initialize theta at 1
t<-1

#x is a vector of the x's from the data and y is a vector of the y's from the data
#iterate 20 times through the N-R algorithm
#u is the first derivative
#v is the second derivative
for(i in 1:20) {
  u <- -sum(x*(y-exp(t*x))*exp(t*x))
  v <- sum(x^2*exp(t*x)*(2*exp(t*x)-y))
  t <- t-u/v
}
```

(2)

Using the code above I obtained $\hat{\theta} \approx 0.99025$

(3)

To find an approximate variance for $\hat{\theta}$ I first approximated $var(y_i)$ using $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-1}$ where $\hat{y}$ is the vector of fitted y's using $\hat{\theta}$

I used R to calculate this:

```r
#calc the fitted values
yh <- exp(t*x)

#calc the residuals
eh <- y-yh

#calc the var and st dev
varhat <- sum(eh^2)/(99)
stdev <- sqrt(varhat)
```

I then ran 1000 parametric bootstrap samples using $\hat{\sigma}^2 \approx 1.01036$ and the N-R algorithm from (1) to get 1000 $\hat{\theta}^*$'s in R:

```r
ttt <- 0                            #create an empty vector to gather bootstrap thetas
for(j in 1:1000)  {                 #run through 1000 bootstrap samples
  tt<-1                             #initialize theta_hat at 1 for each sample
  yy <- exp(t*x)+stdev*rnorm(100)   #generate 100 y's using estimators for theta and variance

  for(i in 2:20) {                   #use N-R algorithm with new y's to generate a new theta_hat*
    uu <- -sum((yy-exp(tt*x))*x*exp(tt*x))
    vv <- sum(x^2*exp(tt*x)*(2*exp(tt*x)-yy))
    tt <- tt-uu/vv
  }

ttt[j]<- tt                         #collect the current iteration's theta_hat*
}

var(ttt)                            #find the variance of the bootstrapped theta_hat*'s
```

(4)

Per the last line of my code above I found $Var(\hat{\theta}) \approx 0.0001141921$

(5)

To find a 95% confidence interval for $\log \hat{y}$ we can use the variance from **(4)** to find the std. dev. of $\hat{\theta}\bar{x}$, I did this in R:

```
xbar <- mean(x)                 #find mean of x values
sd <- sqrt((xbar^2)*var(ttt))   #use the mean to find std dev of log(y)
```

And got $sd = \sqrt{Var(\hat{\theta}\bar{x})} = \sqrt{\bar{x}^2 Var(\hat{\theta})} \approx \sqrt{0.0001141921 \times 0.0072053} \approx 0.0008636511$

```
t*xbar + (1.96*sd)              #upper bound
t*xbar - (1.96*sd)              #lower bound
```

The 95% confidence interval came out to be $(-0.08574931, -0.0823638)$

Problem 2

(1)

To find $l'(\theta)$ I used the chain rule with $f(\theta) = \frac{1}{2}\sum_{i=1}^{n}(x)^2$ and $g(\theta) = y_i - m(\theta, x_i)$

We get $f'(\theta) = \sum_{i=1}^{n}(x)$ and $g'(\theta) = m'(\theta, x_i)$ where $m'(\theta, x_i)$ is the derivative of $m(\theta, x_i)$

This gives us $l'(\theta) = \sum_{i=1}^{n}(y_i - m(\theta, x_i))m'(\theta, x_i)$

(2)

$$
\begin{aligned}
E[l'(\theta)] &= E\left[\sum_{i=1}^{n}(y_i - m(\theta, x_i))m'(\theta, x_i)\right] \\
&= \sum_{i=1}^{n}(E[y_i] - m(\theta, x_i))m'(\theta, x_i) \\
&= \sum_{i=1}^{n}(m(\theta, x_i) - m(\theta, x_i))m'(\theta, x_i) \\
&= \sum_{i=1}^{n}(0)m'(\theta, x_i) \\
&= 0
\end{aligned}
$$

(3)

Since $\hat{\theta}$ is the value that minimizes $l(\theta)$, $l'(\hat{\theta}) = 0$
So we have:

$$
\begin{aligned}
l'(\hat{\theta}) &\approx l'(\theta) + (\hat{\theta} - \theta)l''(\theta) \\
0 &\approx l'(\theta) + (\hat{\theta} - \theta)l''(\theta) \\
\hat{\theta} - \theta &\approx -\frac{l'(\theta)}{l''(\theta)}
\end{aligned}
$$

(4)

To find the mean of the approximate normal distribution for $\hat{\theta}$ we find $E[\hat{\theta}]$:

$$E[\hat{\theta}] \approx E\left[\theta - \frac{l'(\theta)}{l''(\theta)}\right] = E[\theta] - E\left[\frac{l'(\theta)}{l''(\theta)}\right] = E[\theta] - 0 = \theta$$

So the mean is θ

(5)

Similarly to find the variance for $\hat{\theta}$ we find:

$$Var(\hat{\theta}) = Var\left(\theta - \frac{l'(\theta)}{l''(\theta)}\right) = Var(\theta) - Var\left(\frac{l'(\theta)}{l''(\theta)}\right) \approx 0 + \frac{Var(l'(\theta))}{(l''(\theta))^2}$$

We see that:

$$\begin{aligned} Var(l'(\theta)) &= Var\big(\sum_{i=1}^{n}(y_i - m(\theta, x_i))m'(\theta, x_i)\big) \\ &\approx \sum_{i=1}^{n} Var[(y_i - m(\theta, x_i))m'(\theta, x_i)] \\ &\approx \sigma^2 \sum_{i=1}^{n}(m'(\theta, x_i))^2 \end{aligned}$$

So $Var(\hat{\theta}) \approx \sigma^2 \frac{\sum_{i=1}^{n}(m'(\theta, x_i))^2}{(l''(\theta))^2}$