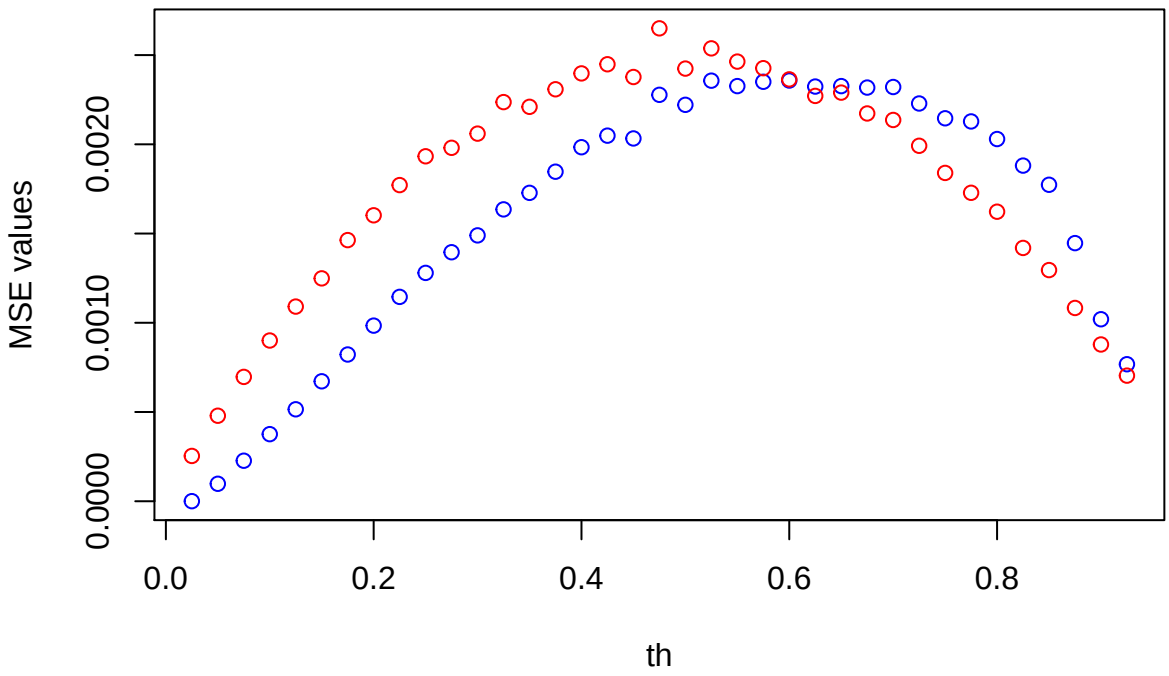
Lab Week 8

Donald Elrod

May 10, 2017

Question 1

```
th = (1:37)/40
p.th = sqrt( (40*th-1)/36)
th.hat.3=th.hat.2=MSE3=MSE2=0
for (i in 1:37) {
  for (j in 1:5000) {
    pop = rep(c(p.th[i], 0.5), c(9,1))
    random.p = sample(pop, size=100, replace=T)
    samp = rbinom(100, 2, random.p)
    p.3 = (mean(samp)-0.1)/1.8;
    p.3 = min( max(0, p.3), 1)
    th.hat.3[j] = (36*p.3^2 + 1)/40;
    th.hat.2[j] = mean(samp==2)
  }
  mth3 = (th.hat.3 - th[i])^2
  mth2 = (th.hat.2 - th[i])^2
  MSE3[i] = mean(mth3)
  MSE2[i] = mean(mth2)
}

ymax = max(MSE3, MSE2)

plot(th, MSE3, col='blue', ylim=c(0, ymax), xlab = 'th', ylab = 'MSE values')
points(th, MSE2, col='red')
```




```
summary(MSE2)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0002538 0.0012950 0.0019800 0.0017770 0.0023080 0.0026500
```

```
summary(MSE3)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 3.712e-07 1.020e-03 1.847e-03 1.591e-03 2.229e-03 2.357e-03
```

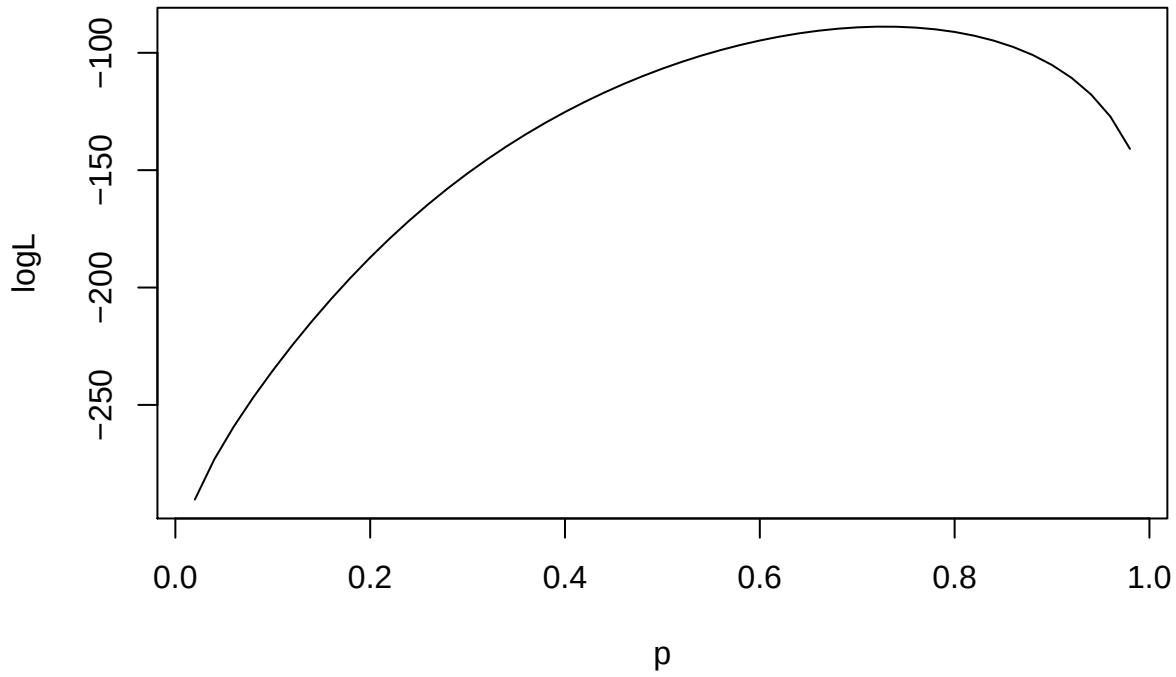
Question 2

θ_2 seems a better fit for the data, as it seem more normally distributed than θ_3, however θ_3 is a better model in the range of ~ 0.6

Question 3

```
pop = rep(c(0.7, 0.5), c(9,1))
random.p = sample(pop, size=100, replace=T)
samp = rbinom(100, 2, random.p)
p = (1:49)/50

binom.mix.logL=function(p,data) {
```

```
  probdist=0.9*dbinom(data,2,p)+0.1*dbinom(data,2,0.5)
  sum(log(probdist))
}

logL=0

for(i in 1:49){
  logL[i]=binom.mix.logL(p[i],samp)
}

plot(p, logL, type='l')
```



Question 4

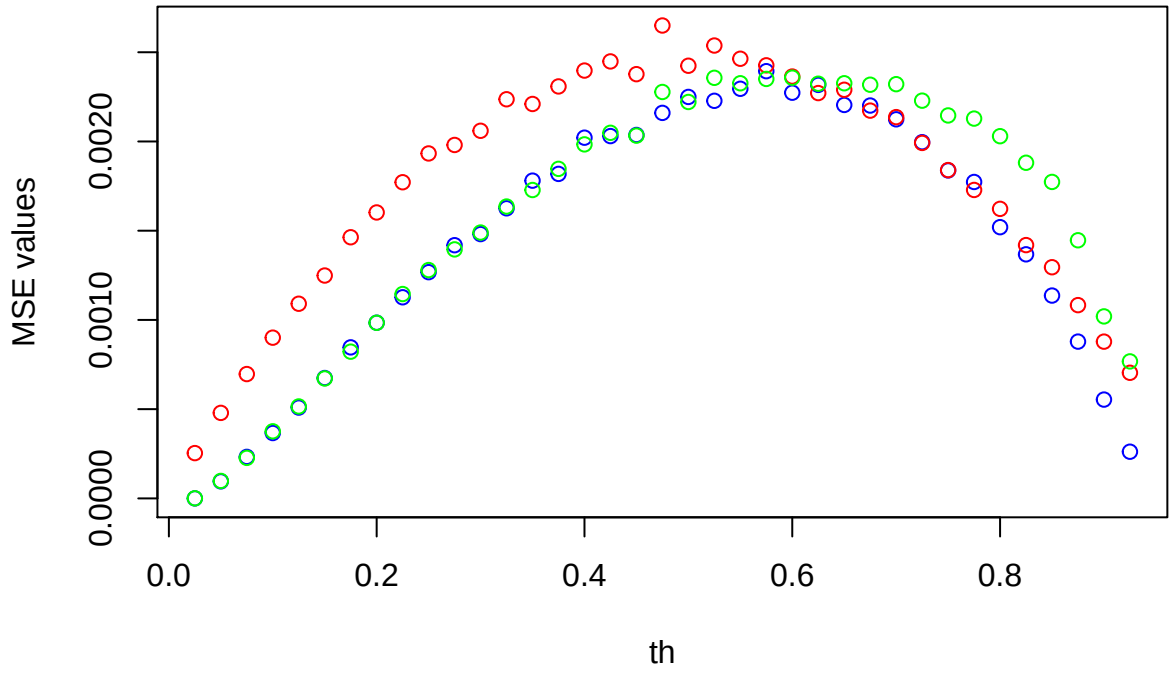
```
fit = optimise(binom.mix.logL, c(0,1), data=samp, maximum=T)
fit$max
```

```
## [1] 0.728424
```

Question 5

```
th = (1:37)/40
p.th = sqrt( (40*th-1)/36)
th.hat.4=MSE4=0
for (i in 1:37) {
  for (j in 1:5000) {
    pop = rep(c(p.th[i], 0.5), c(9,1))
    random.p = sample(pop, size=100, replace=T)
    samp = rbinom(100, 2, random.p)
    p.4 = optimise(binom.mix.logL, c(0,1), data=samp,maximum=T)$max
    th.hat.4[j] = (36*p.4^2 + 1)/40;
  }
  mth4 = (th.hat.4 - th[i])^2
  MSE4[i] = mean(mth4)
}
```

```
ymax = max(MSE2, MSE3, MSE4)

plot(th, MSE4, col='blue', ylim=c(0, ymax), xlab = 'th', ylab = 'MSE values')
points(th, MSE2, col='red')
points(th, MSE3, col='green')
```



Question 6

θ_2 seems to have a better fit for the simulation when the range is ~0.5, whereas θ_3 and θ_4 appear to be more accurate in the range of ~0.6-7, with θ_3 being closer to 0.7 and θ_4 being closer to 0.6