STATS205HW1

Yuhshang

1/8/2020

Problem 1

(1)

$$p(\theta|x=3) = \frac{p(x=3|\theta)p(\theta)}{p(x=3)}$$

```
ptheta    = c(.05,.05,.8,.05,.05);
theta     = c(0.3,0.4,0.5,0.6,0.7);
px3theta = 5*4*theta^3*(1-theta)^2*ptheta;
px3       = 5*4*sum(theta^3*(1-theta)^2*ptheta);
pthetax3 = px3theta/px3
```

Then, the posterior P(0.3 0.4 0.5 0.6 0.7) is

```
print(pthetax3)
```

```
## [1] 0.02198770 0.03829151 0.83097889 0.05743726 0.05130464
```

(2)

$$E(\theta) = \sum \theta * p(\theta|x=3)$$

```
Ethetax3 = sum(pthetax3*theta);
print(Ethetax3)
```

```
## [1] 0.507778
```

(3)

$$P(\theta|x=6) = \frac{P(x=6|\theta) * P(\theta)}{P(x=6)})$$

$$P(\theta|x=6) = \frac{P(x=6|\theta) * P(\theta|x=3)}{\sum_{\theta*} P(x=6|\theta^*) * P(\theta^*|x=3)})$$

Thus, the new inference is

```
px6      = choose(7,6) *sum(theta^6*(1-theta)*pthetax3);
pthetax6 = choose(7,6) *theta^6*(1-theta)*pthetax3/px6;
pthetax6
```

```
## [1] 0.001183573 0.009926661 0.684809113 0.113070867 0.191009787
```

Problem 2

(1)

The prior probability that $\theta > 0.5$ is:

```
theta2  = c(0,0.125,0.25,0.375,.5,.625,.750,0.875,1);
ptheta2 = c(.001,.001,.95,.008,.008,.008,.008,.008,.008);
sum(ptheta2[6:9])
```

```
## [1] 0.032
```

(2)

$$P(x = 7) = \sum_{\theta} P(x = 7|\theta) * P(\theta)$$

```
px7     = sum(dbinom(7,size=10,prob = theta2)*ptheta2);
print(px7)
```

```
## [1] 0.008741856
```

(3)

```
px6_2   = sum(dbinom(6,size=10,prob=theta2)*ptheta2);
ptheta2x6 = dbinom(6,size=10,prob=theta2)*ptheta2/px6_2;
print(sum(ptheta2x6[6:9]))
```

```
## [1] 0.1579494
```

Problem 3

(1)

I will consider binomial distribution for the likelihood function $\sim Bin(n, \theta)$

(2)

I will choose beta distribution since I have a probability of 20% success for the pior information.

(3)

$$P(\theta|x = 14) = \frac{P(x = 14|\theta) * P(\theta)}{P(x = 14)}$$

$$P(\theta|x = 14) = \frac{P(x = 14|\theta) * Beta(1, 4)}{P(x = 14)}$$

(4)

$$E[\theta] = \frac{a}{a + b} = 1/5$$

$$E[\theta|x = 14] = \frac{a + x}{a + b + m} = \frac{1 + 14}{5 + 30} = 3/7$$

(5)

$$P(\theta > 0.25|x = 14) = pbeta(1, 15, 20) - pbeta(0.25, 15, 20)$$

```
pthetal25 = pbeta(1,15,20)-pbeta(0.25,15,20);
print(pthetal25)
```

```
## [1] 0.9883369
```

(6)

The posterior theta gets larger then the pior after winter 2020.

(7)

$$P(x = 10|y) = \int_{\theta} Bin(10, \theta) * Beta(15, 20) d\theta$$

By Monte-Carlo, the probability that number of success s greater than 5 is:

```
post_samples = rbeta(10000,15,20);
pred_samples = sapply(post_samples,rbinom,size=20,n=1);
sum(pred_samples>5)/length(pred_samples)
```

```
## [1] 0.8648
```

(8)

I will choose Beta(15,20) as the new pior which is the posterior of last inference.

By formula in class notes, y is the new success number 10, m is the total new number 30, a = 15, b =20

$$\theta|x \ Beta(a + y, b + m - y)$$

$$E[\theta|x] = (a + y)/(a + b + m) = 25/65 = 5/13$$

By (8), y = 10 , m = 30, thus new expection.

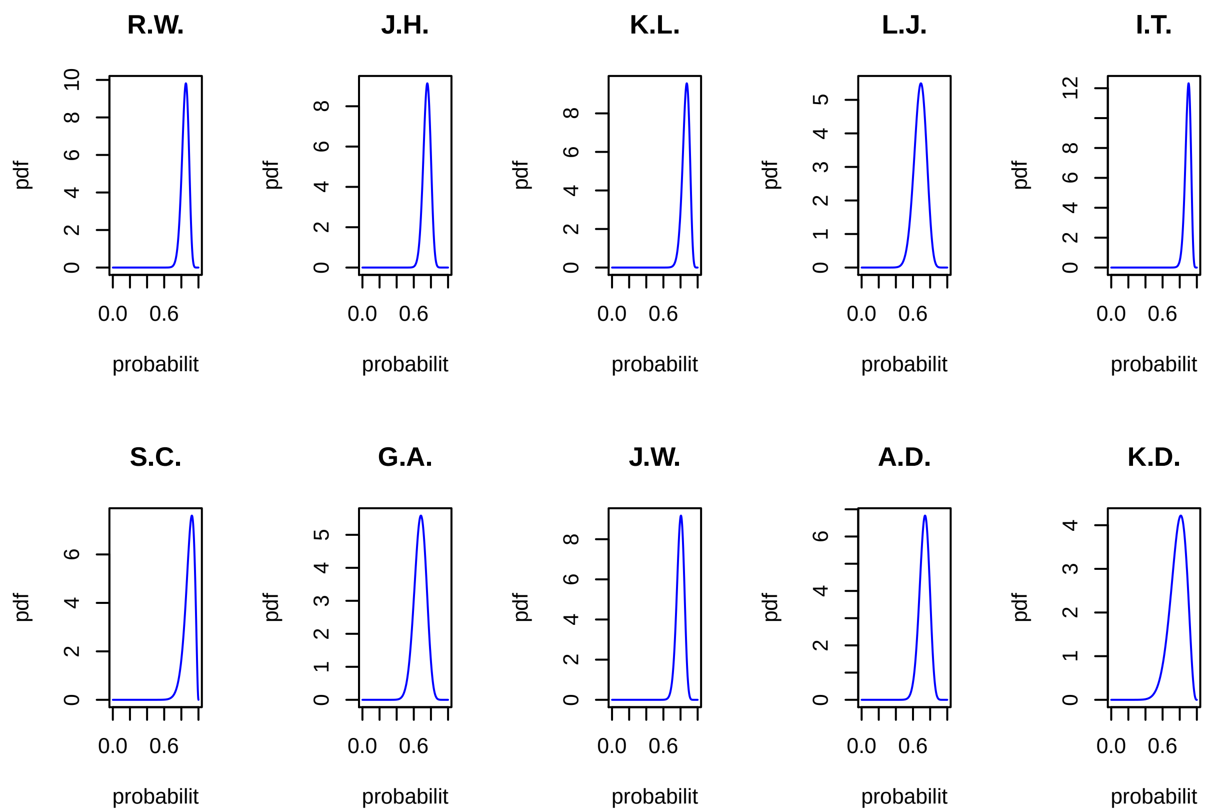
Problem 4

1

The likilihood I will use binomial distribtion with n = Clutch attempts and θ = Clutch makes / Clutch attempts. And the prior I will use uniform distribution which is beta(1,1).

2

```
makes = c(64,72,55,27,75,24,28,66,40,13);
attem = c(75,95,63,39,83,26,41,82,54,16);
names = c('R.W.','J.H.','K.L.','L.J.','I.T.','S.C.','G.A.','J.W.','A.D.','K.D.')
par(mfrow=c(2,5));
x_lst = seq(0,1,0.0025);
for(i in 1:10){
y_lst = dbeta(x_lst,(1+makes[i]),(1+attem[i]-makes[i]));
plot(x_lst,y_lst,main=names[i],xlab = 'probabilit',ylab = 'pdf',type="l",col = 'blue')
plot.new
}
```



3

```
a =1;
b =1;
data0      = matrix(NA,11,7);
post_mean = (a+makes)/(a+b+attem);
CI     = c(.05,.25,.5,.75,.95);
data0[1,] = c('name quantile',CI,'mean');
for (i in 2:11){
  i = i-1;
  data0[i+1,] = c(names[i],qbeta(CI,a+makes[i],1+attem[i]-makes[i]),post_mean[i])
}
table0 = data.frame(data0);
table0
```

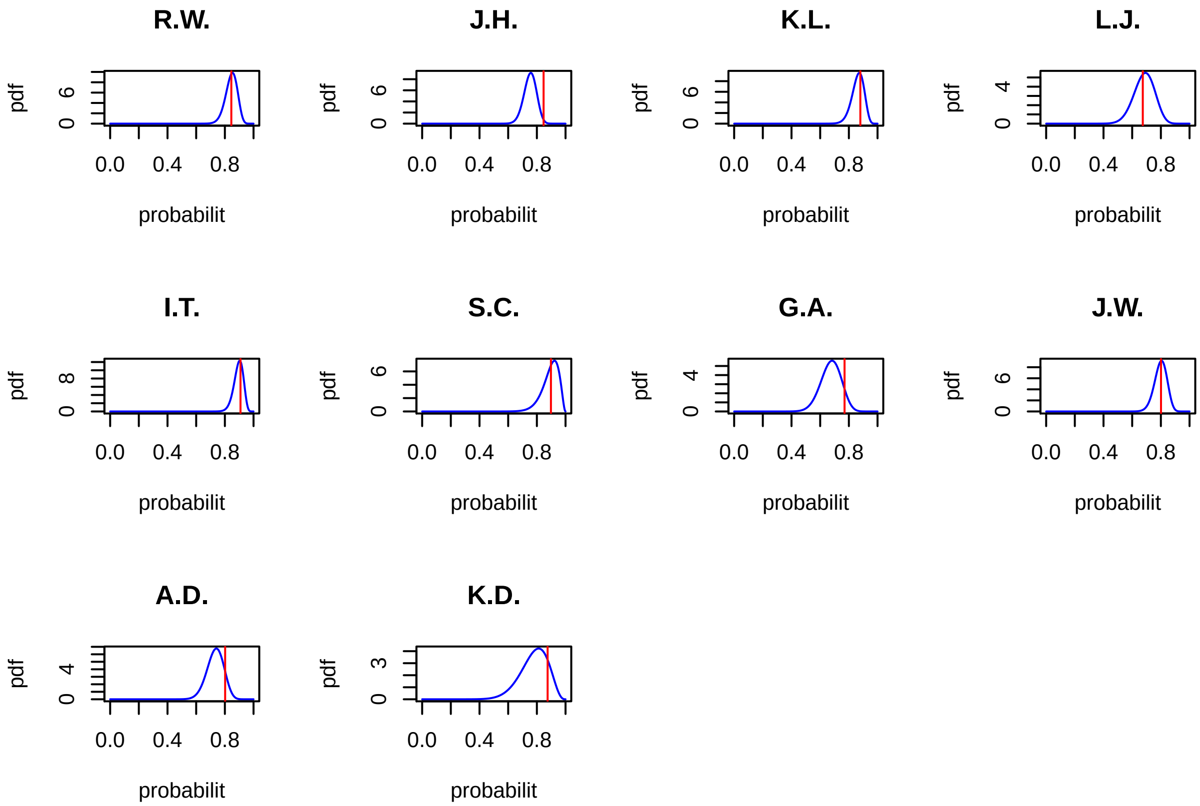
```
##                X1                X2                X3                X4
## 1  name quantile              0.05              0.25               0.5
## 2            R.W. 0.771815740875448 0.818080157614331 0.847142177379137
## 3            J.H. 0.677973913016999 0.724003527064116 0.754318868024149
## 4            K.L. 0.785744329691379 0.834945611763296 0.865253244754848
## 5            L.J. 0.559720262234269 0.635494861183767 0.685927146113428
## 6            I.T. 0.834716421051674 0.873561159044628 0.897208109472869
## 7            S.C. 0.784700038857546 0.859906054307051 0.902187233911541
## 8            G.A. 0.553564929507167 0.627647745441107 0.677145498234897
## 9            J.W. 0.722054842518034 0.769426690492551 0.799988461656181
## 10           A.D. 0.631022116805061 0.693774078054947 0.734922651225345
## 11           K.D. 0.604358566555179 0.717925704606302 0.788214999102497
##                   X5                X6                X7
```

```
## 1                0.75              0.95              mean
## 2  0.873456089170816 0.906289153453488 0.844155844155844
## 3   0.78303641103931 0.821231411412196 0.752577319587629
## 4  0.892133999275377 0.924630991222677 0.861538461538462
## 5  0.733576615013464 0.795865904203997 0.682926829268293
## 6  0.918011243793005 0.942956307701054 0.894117647058824
## 7  0.935680174612185 0.969021606765964 0.892857142857143
## 8  0.724116420645216 0.785941833611734 0.674418604651163
## 9  0.828372966175675 0.865086922075984 0.797619047619048
## 10 0.773505595922638 0.823758110241969 0.732142857142857
## 11 0.848559659157339 0.915354897969421 0.777777777777778
```

4

By ploting the overall propotion on the beta distribution.

```r
op = c(.845,.847,.880,.674,.909,.898,.77,.801,.802,.875);
makes = c(64,72,55,27,75,24,28,66,40,13);
attem = c(75,95,63,39,83,26,41,82,54,16);
names = c('R.W.','J.H.','K.L.','L.J.','I.T.','S.C.','G.A.','J.W.','A.D.','K.D.')
par(mfrow=c(3,4));
x_lst = seq(0,1,0.0025);
for(i in 1:10){
y_lst = dbeta(x_lst,(1+makes[i]),(1+attem[i]-makes[i]));
plot(x_lst,y_lst,main=names[i],xlab = 'probabilit',ylab = 'pdf',type="l",col = 'blue')
abline(v=op[i],col="red")
plot.new
}
```




```r
lar = c(0)
for(i in 1:10){
  lar[i] = 1-pbeta(op[i],(1+makes[i]),(1+attem[i]-makes[i]))
}

data1     = matrix(NA,10,2);

for (i in 1:10){
  data1[i,] = c(names[i],lar[i])
}
table1 = data.frame(data1)
table1
```

```
##      X1                  X2
## 1  R.W.   0.520712117345203
## 2  J.H. 0.00902080105142355
## 3  K.L.   0.359755854305227
## 4  L.J.   0.564554731776409
## 5  I.T.   0.355375348458648
## 6  S.C.   0.529322852225301
## 7  G.A.  0.0833432944114901
## 8  J.W.   0.490763382615694
## 9  A.D.    0.11333610410631
## 10 K.D.   0.154254131745749
```

From the previous table, We find that R.W. L.J. S.C. all the more than half clutch numbers than than their overal precentage they should do better based on distribtion. Thus, they are different. They preform not as good as they should be. On the other hand J.H. G.A. A.D. and K.D.;s overall propotions too high than they should be based. So, they perform realy good in 2006-2007 sesaon.