STAT_205_HW1

HONGLEI REN

1/17/2020

Problem 1

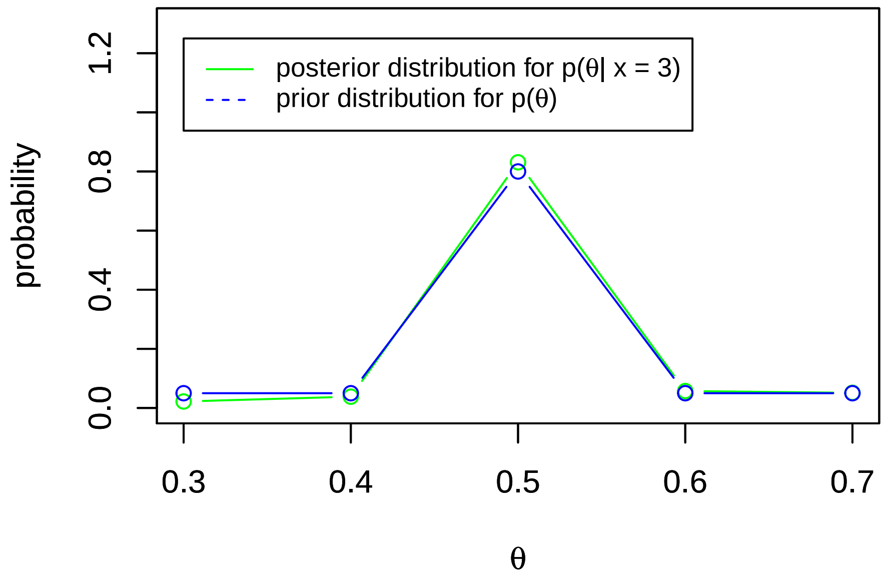
1.1 Posterior Distribution

```
n = 5
y = 3
theta = c(0.3, 0.4, 0.5, 0.6, 0.7)
prior_theta = c(0.05, 0.05, 0.8, 0.05, 0.05) #prior distribution
```

```
likelihood_data = dbinom(y, n, theta) # likelihood of data
posterior_theta = (likelihood_data * prior_theta) / sum(likelihood_data * prior_theta) #posterior
names(posterior_theta) = theta
round(posterior_theta, 2)
```

```
##  0.3  0.4  0.5  0.6  0.7
## 0.02 0.04 0.83 0.06 0.05
```

```
plot(theta, posterior_theta,col="green", type="b",
     xlab = expression(theta), ylab = "probability", ylim=range(c(0,1.3)))
par(new=TRUE)
plot(theta, prior_theta,col="blue", type="b",
     xlab = expression(theta), ylab = "probability", ylim=range(c(0,1.3)))
legend(0.3, 1.25, legend=c(
  expression(paste("posterior distribution for p(", theta, "| x = 3)")),
  expression(paste("prior distribution for p(", theta, ")"))),
       col=c("green", "blue"), lty=1:2, cex=0.8)
```



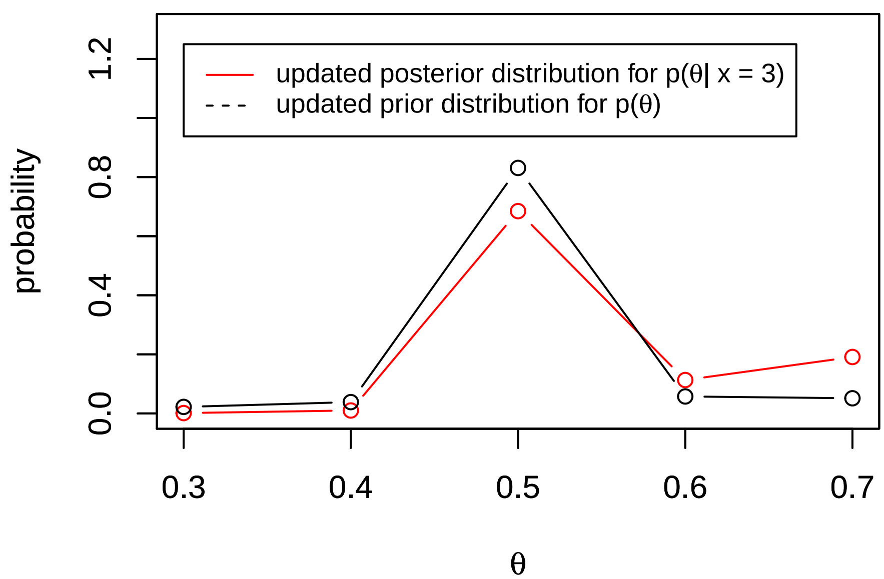
1.2 Posterior Mean

```
expectation_of_theta = sum(theta * posterior_theta)
```

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

1.3 Update Posterior by New Data

```
n = 7
y = 6
theta = c(0.3, 0.4, 0.5, 0.6, 0.7)
prior_theta = posterior_theta
likelihood_data = dbinom(y, n, theta) # likelihood of data
posterior_theta = (likelihood_data * prior_theta) / sum(likelihood_data * prior_theta) #posterior
```



Problem 2

2.1 Prior Probability

```
theta = c(0, .125, .250, .375, .500, .625, .750, .875, 1)
prior_theta = c(.001, .001, .950, .008, .008, .008, .008, .008, .008)
prior_prob = sum(prior_theta[theta > 0.5])
```

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

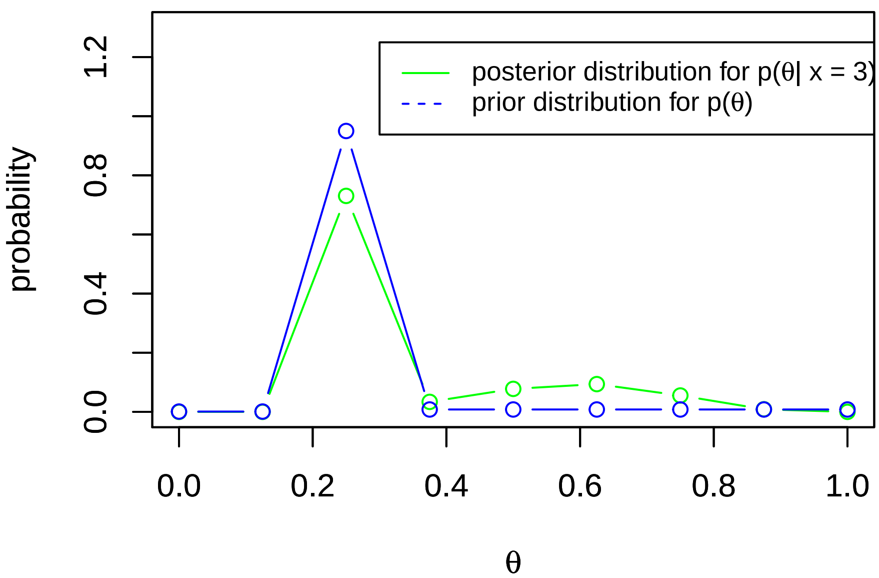
2.2 Prior Predictive

```
n = 10
y = 7
prior_predictive = sum(dbinom(y, n, theta))
```

```
## [1] 0.7289292
```

0.4 Posterior Probability

```
n = 10
y = 6
likelihood_data = dbinom(y, n, theta) # likelihood of data
posterior_theta = (likelihood_data * prior_theta) / sum(likelihood_data * prior_theta) #posterior
```



The posterior probability that Bob has ESP ability is:

```
posterior_prob = sum(posterior_theta[theta > 0.5])
posterior_prob
```

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

Problem 3

3.1 Likelihood Function

The likelihood function I considered for this problem is $Y \sim Bin(n, \theta)$, where Y is the number of students that spells correctly, θ is the success rate of spelling.

3.2 Prior Distribution

Based on the prior of 20% success rate with a sample size 5, the prior could be $\theta \sim Beta(a = 2, b = 3)$.

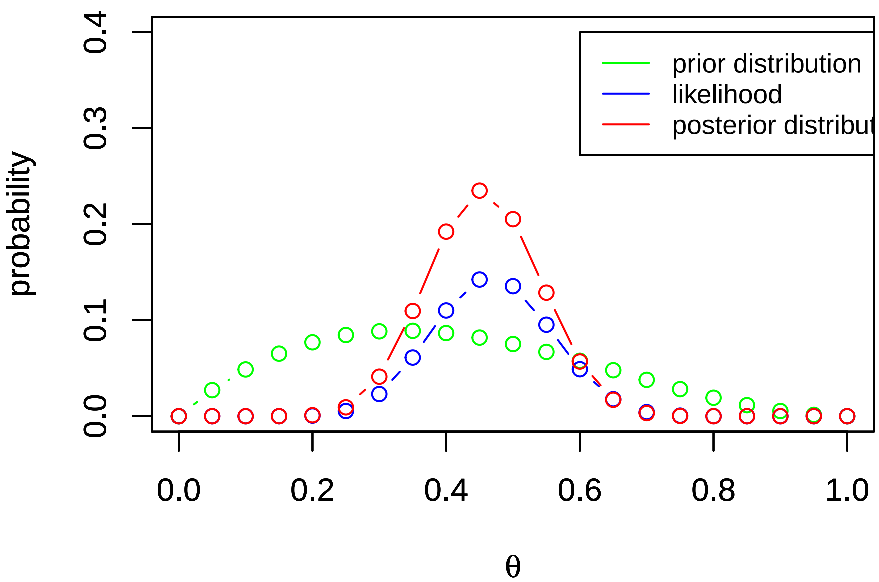
3.3 Posterior Distribution

The mathematical expression of the posterior distribution is:

$$P(\theta|Y) \sim \frac{P(Y|\theta)P(\theta)}{P(Y)} = c\theta^{y+a-1}(1-\theta)^{n-y+b-1} = Beta(y + a, n - y + b)$$

, where $n = 30, y = 14, a = 2, b = 3$.

```
a = 2; b = 3; n = 30; y = 14
theta = seq(0, 1, by=0.05)
prior_theta = dbeta(theta, a, b)
likelihood_data = dbinom(y, n, theta)
posterior_theta = (likelihood_data * prior_theta) / sum(likelihood_data * prior_theta)
```



3.4 Expectation

The posterior expectation is:

```
## [1] 0.4571429
```

And the prior expectation is:

```
## [1] 0.401
```

Based on the posterior expectation compared with prior expectation, I think the effectiveness of his training practices has improved, because the average sucess rate of spelling was increased.

3.5 Posterior Probablity that probablity of sucess in spelling is greater than 25%

The posterior probablity that probablity of sucess in spelling is greater than 25% is:

```
## [1] 0.98966
```

3.6 Difference Interpretion

The difference between the posterior and prior is a reflection of the change in the effectiveness of training practice or any related factors.

3.7 Posterior Predictive Probablity

The Posterior Predictive Probablity is:

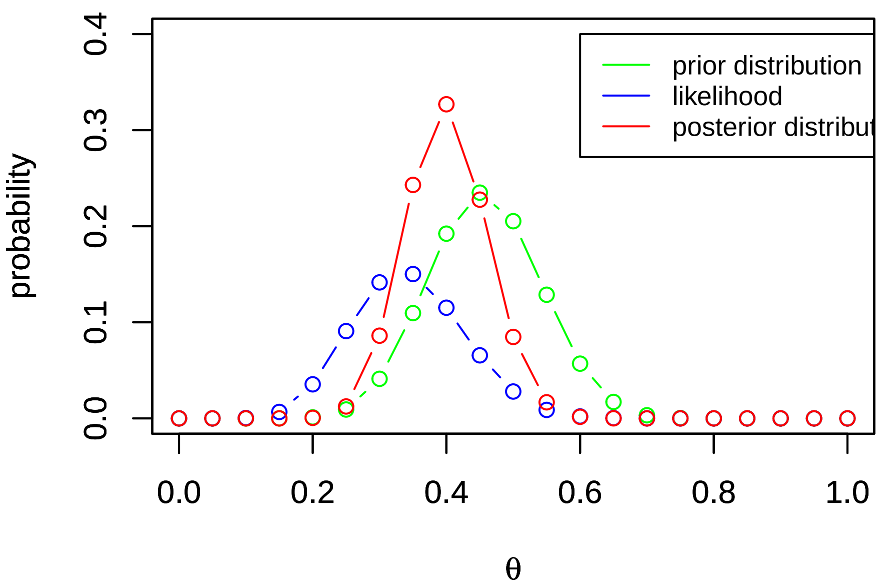
```
A = y + a
B = n - y + b
theta_hat = A / (A + B)
n = 20
likelihood = sum(dbinom(5:n, n, theta_hat))
likelihood
```

```
## [1] 0.9838477
```

3.8

I choose $\theta \sim Beta(A = 16, B = 19)$ according to the previous data. The posterior is shown in the following figure, and the posterior expectation is:

```
## [1] 0.4
```



Problem 4

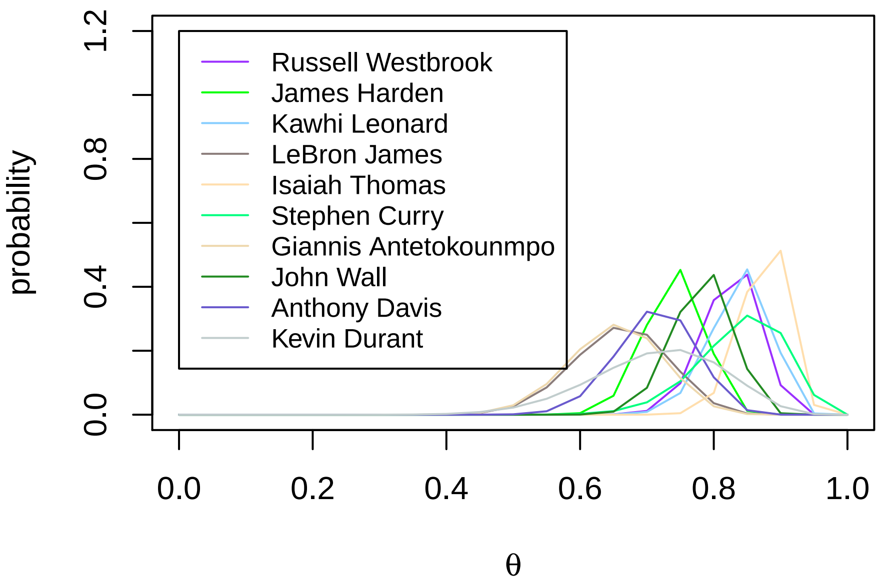
4.1 Model Description

Because I don't have any knowledge about the clutch success probabilty, I assumed a uniform prior $\theta = \frac{1}{1001}$ for (0, 0.001, 0.002, ..., 0.999, 1.0). The likelihood would be an Binomial with θ as parameter.

4.2 Posterior

```
theta = seq(0, 1, by=0.05)
prior_theta = dbeta(theta, 1, 1)
data_summary = matrix(nrow =nsample, ncol = 6)
plot(1,type='n', xlab = expression(theta), ylab = "probability",xlim = range(c(0, 1.0)), ylim=range(c(0
for (idx in seq(1, nsample))
{
  y = df$cmake[idx]
  n = df$cattempts[idx]
  A = y + a;
  B = n - y + b;
  theta_hat = A / (A + B)
  theta_star = rbeta(S, A, B)
  posterior_theta = dbeta(theta, A, B)
  posterior_theta = posterior_theta/sum(posterior_theta)
  lines(theta, posterior_theta, col=colors[idx])
  qtl = quantile(theta_star, probs= probs)
  mn = mean(theta_star)
  data_summary[idx, ] = c(mn, qtl)
  theta_cdf = ecdf(theta_star)
  prob_greater_than_overall = c(prob_greater_than_overall, theta_cdf(1.) - theta_cdf(df$overall_proporti
}

legend(0., 1.2, legend=as.character(df$players),
       col=colors, lty=1:1, cex=0.8)
```



4.3 Summarize the Posterior

```r
library(knitr)
dt = as.data.frame(data_summary, as.character(df$players))
colnames(dt) = c("mean", "5% percentile", "25% percentile", "50% percentile", "75% percentile", "95% pe
kable(dt, caption="Summary for posterior")
```

Table 1: Summary for posterior

	mean	5% percentile	25% percentile	50% percentile	75% percentile	95% percentile
Russell Westbrook	0.8250490	0.7509871	0.7981135	0.8276436	0.8549652	0.8894384
James Harden	0.7400113	0.6654576	0.7110512	0.7416748	0.7706129	0.8095037
Kawhi Leonard	0.8384012	0.7604694	0.8103216	0.8418792	0.8700916	0.9048964
LeBron James	0.6591138	0.5393065	0.6120603	0.6613414	0.7088129	0.7713946
Isaiah Thomas	0.8748090	0.8123005	0.8527693	0.8775814	0.8999646	0.9273524
Stephen Curry	0.8384992	0.7204791	0.7986580	0.8455723	0.8860091	0.9315798
Giannis Antetokounmpo	0.6522310	0.5341731	0.6064515	0.6546000	0.7005752	0.7624577
John Wall	0.7816902	0.7055077	0.7532073	0.7837841	0.8126228	0.8502535
Anthony Davis	0.7119948	0.6110962	0.6736897	0.7145377	0.7530782	0.8038133
Kevin Durant	0.7143121	0.5448472	0.6516074	0.7211623	0.7840122	0.8599875

4.4 Probability clutch > overall proportion

The probability clutch > overall proportion is summarized as follows:

```r
ds = data.frame("players" = df$players, "Prob" = prob_greater_than_overall)

kable(ds, caption="Table for probability clutch > overall proportion")
```

Table 2: Table for probability clutch > overall proportion

players	Prob
Russell Westbrook	0.33687
James Harden	0.00328
Kawhi Leonard	0.17656
LeBron James	0.42958
Isaiah Thomas	0.16571
Stephen Curry	0.18376
Giannis Antetokounmpo	0.03868
John Wall	0.34575
Anthony Davis	0.05399
Kevin Durant	0.03069

From the table above, we can see all players have a lower probility for clutch than overall proportion, and James Harden, Giannis, Anthony Davis, Durant has extremely different clutch percentage.