STA 365 Home Work #1 Ruike Xu.

Problem 1. (15 points)

(a) Provide an example of a sequence of random variables, $X_1, ..., X_n$, which are exchangeable but not iid.

(b) Let $X_1, ...X_n$ be a sequence of continuous exchangeable random variables. Prove that $X_1, ...X_n$ are identically distributed.

(a)

Suppose that we have an urn containing 1 red balls and 3 green balls.
Balls are drawn out of the box one at a time without replacement.
Let

$$X_i = \begin{cases} 1 & \text{if the ball drawn is red} \\ 0 & \text{otherwise} \end{cases}$$

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) > 0 \text{ if } \exists X_i = 1$$

$$P(X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0) = \frac{1}{4} \times 1 \times 1 = \frac{1}{4}$$

$$P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0) = \frac{\cancel{3}}{4} \times \frac{\cancel{2}}{\cancel{3}} \times \frac{1}{\cancel{2}} \times 1 = \frac{1}{4}$$

$$P(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0) = \frac{\cancel{3}}{4} \times \frac{1}{\cancel{3}} \times 1 \times 1 = \frac{1}{4}$$

Thus the random variable X_1, X_2, X_3, X_4 are exchafeable

$$P(X_1 = 1) = \frac{1}{4}$$

$$P(X_2 = 1 \mid X_1 = 0) = \frac{1}{3}$$

Therefore, if we don't get the red ball on the first draw, we have a higher chance of getting the red ball in the next draws. There exists dependency between the outcomes of the random variables.

Thus, this sequence of random variables is exchangeable but not iid.

(b) Suppose that $X_1, X_2, \cdots, X_n$ are Continuous exchangeable random Variables

Let i and j be $\in 1, \cdots n$

$$f_{X_i}(x_i) = \int\int\cdots\int f_{X_1,X_2,\cdots,X_n}(x_1,\cdots,x_i,\cdots,x_j,\cdots x_n)\, dx_1, \cdots dx_j \cdots dx_n$$

By using exchangeable Property of the Sequence, We Can exchange i and j.

$$f_{X_j}(x_i) = \int\int\cdots\int f_{X_1,X_2,\cdots,X_n}(x_1,\cdots,x_j,\cdots,x_i,\cdots x_n)\, dx_1, \cdots dx_j, \cdots dx_n$$

Even if we exchange i and j, we would still hold x_i in the integration.

Since the integration of marginal distribution achieve the same result,

the two random Variables X_i and X_j are identically distributed.

Therefore, for two arbitrary random Variable X_i and X_j in the Sequence, their distribution are identical to each other.

Thus, if $X_1, \cdots X_n$ is a Continuous exchangeable Sequence of random Variables, they are identical distributed.

Problem 2. (20 points)

You are tasked with predicting the occurrence of a binary event E. That is, $\Omega = E \cup \overline{E}$, and $E \cap \overline{E} = \emptyset$. Let p and $1-p$ denote your true probabilities for E and $\overline{E}$, respectively. Let r and $1-r$ denote your reported probabilities for E and $\overline{E}$, respectively. And suppose that your payoff is given by

$$s(r) = \begin{cases} -\log(r), & E \\ -\log(1-r), & \overline{E} \end{cases}$$

(a) Recall that $s(r)$ is strictly proper if $\mathrm{E}_p s(p) > \mathrm{E}_p s(r)\ \forall r \neq p$. Prove that $s(r)$ is (or is not) strictly proper.

(b) Let $r = p$ and plot $\mathrm{E}_p[s(p)]$ (using a base 2 log) in R. This function is also known as the Shannon information entropy corresponding to the Bernoulli random variable X where E corresponds to $X = 1$ and $\overline{E}$ corresponds to $X = 0$.

(c) If you want to pick a prior probability for E, i.e., $p(E)$, that maximizes Shannon information entropy, which value of p would you select? Hint: find $\arg\max_{p\in[0,1]} \mathrm{E}_p[s(p)]$.

(d) Argue in one paragraph why (or why not) using the measure of Shannon information entropy to select a prior probability is a good/bad idea.

(a) $E_p[s(r)] = P(-\log(r)) + (1-P)(-\log(1-r))$

$= -\log(r)\cdot P - \log(1-r) + P\log(1-r)$

$= -\log(r)\cdot P - (1-P)\log(1-r)$

$E_p[s(P)] = P(-\log(P)) + (1-P)(-\log(1-P))$

$= -\log(P)P - \log(1-P) + P(\log(1-P)$

$= -\log(P)P - (1-P)(\log(1-P))$

Counter example: Suppose we take $r = 0.3$ & $P = 0.4$ and we use log as base 2 log.

$E_p[s(r)] = -\log_2(0.3)\cdot 0.4 - (0.6)\log_2(0.7)$

$= 1.004$

$E_p[s(P)] = -\log_2(0.4)0.4 - (0.6)(\log_2(0.6))$

$= 0.971$

So $E_p[s(P)] \not> E_p[s(r)]\ \forall r \neq P$

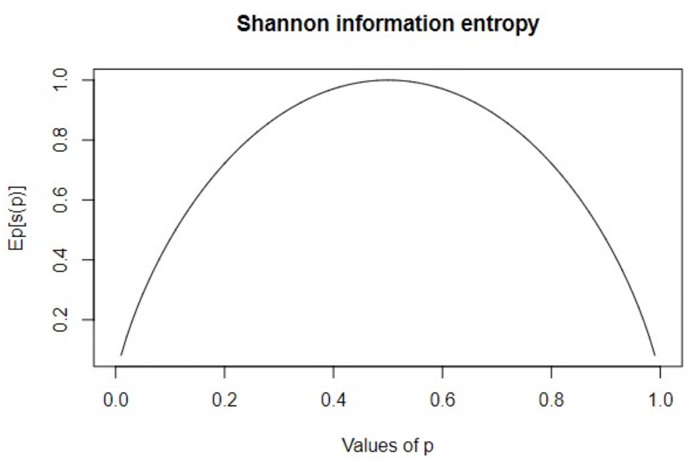
Thus $s(r)$ is not strictly proper.

(b)

Question 2

Part b

```
# Assume that r = p
p <- seq(0, 1, by=0.01)
Ep_Sp <- -log2(p)*p - (1-p)*log2(1-p)
plot(p, Ep_Sp,type="l", xlab="Values of p", ylab="Ep[s(p)]", main ="Shannon information entropy")
```



(c) Want to find $\arg\max_{P\in[0,1]} E_P[S(P)]$, here we use natural log as $\log_{10}$.

$$\arg\max_{P\in[0,1]} -\log(P)\cdot P-(1-P)\log(1-P)$$

$$\frac{d}{dP}\left[-\log(P)\cdot P-(1-P)\log(1-P)\right]=0$$

$$-\left[\frac{d}{dP}(P)\cdot\log(P)+P\cdot\frac{d}{dP}(\log(P))\right]-\left[\frac{d}{dP}(\log(1-P))\cdot(1-P)+\log(1-P)\cdot\frac{d}{dP}(1-P)\right]=0$$

$$-\log(P)-1-\frac{1}{\cancel{1-P}}\cdot(-1)\cdot\cancel{(1-P)}+\log(1-P)=0$$

$$-\log(P)\cancel{-1}\cancel{+1}+\log(1-P)=0$$

$$\log(1-P)-\log(P)=0$$

$$\log(1-P)=\log(P)$$

$$e^{\log(1-P)}=e^{\log(P)}$$

$$1-P=P$$

$$2P=1$$

$$P=\frac{1}{2}$$

$$\frac{d^2}{dP^2}\left(\log(1-P)-\log(P)\right)=\frac{1}{1-P}\times(-1)-\frac{1}{P}$$

$$=-\frac{1}{1-P}-\frac{1}{P}$$

for $P=\frac{1}{2}$, $-\frac{1}{1-\frac{1}{2}}-\frac{1}{\frac{1}{2}}=-1\times 2-1\times 2=-4<0$.

Therefore, when $P=\frac{1}{2}$, $E_P[S(P)]$ achieves the maximum value of the function.

We should choose P as $\frac{1}{2}$.

(d) I think using the measure of Shannon information entropy to select a prior probability is a good idea since the using the maximization of shannon information entropy can assist us to employ minimum assumptions of prior information. Also, the maximization of information entropy can help us maximize the payoffs.

Problem 3. (15 points)

Consider two coins C_1 and C_2, with the following characteristics: $\Pr(\text{heads}|C_1) = 0.7$ and $\Pr(\text{heads}|C_2) = 0.3$. Choose one of the coins at random and imagine spinning it repeatedly. Given that the first two spins from the chosen coin are tails, what is the expectation of the number of additional spins until a head shows up?

Solution: Suppose that we need N additional tails to get a head.

$$E[N|C_1] = \sum_{n=0}^{\infty} n \cdot 0.3^n \cdot 0.7$$
$$= 0.7\left(\sum_{n=0}^{\infty} n \cdot 0.3^n\right)$$
$$= 0.7 \times \frac{0.3}{(0.3-1)^2}$$
$$= \frac{3}{7}$$

$$E[N|C_2] = \sum_{n=0}^{\infty} n \cdot 0.7^n \cdot 0.3$$
$$= 0.3\left(\sum_{n=0}^{\infty} n \cdot 0.7^n\right)$$
$$= 0.3 \times \frac{0.7}{(0.7-1)^2}$$
$$= \frac{7}{3}$$

$$P(\text{tail}, \text{tail}) = P(C_1) \times P(\text{tail}, \text{tail} | C_1) + P(C_2) \times P(\text{tail}, \text{tail} | C_2)$$
$$= \frac{1}{2} \times 0.3^2 + \frac{1}{2} \times 0.7^2$$
$$= \frac{29}{100}$$

$$P(C_1 | \text{tail}, \text{tail}) = \frac{P(\text{tail}, \text{tail} | C_1) P(C_1)}{P(\text{tail}, \text{tail})} = \frac{0.3^2 \times 0.5}{0.29} = \frac{9}{58}$$

$$P(C_2 | \text{tail}, \text{tail}) = \frac{P(\text{tail}, \text{tail} | C_2) P(C_2)}{P(\text{tail}, \text{tail})} = \frac{0.7 \times 0.5}{0.29} = \frac{49}{58}$$

Suppose that X is the number of additional spins until a head shows up.

$$E(X) = P(C_1 | \text{tail}, \text{tail}) \times [E(N|C_1) + 1] + P(C_2 | \text{tail}, \text{tail}) \times [E(N|C_2) + 1]$$
$$= \frac{9}{58} \times \left(\frac{3}{7} + 1\right) + \frac{49}{58} \times \left(\frac{7}{3} + 1\right)$$
$$= 3.038$$

Therefore, the expectation of number of additional spins until a head shows up is 3.038.

Problem 4. (20 points)

Suppose we are going to sample 100 individuals from a county (of size much larger than 100) and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

(a) Assume $Y_1, ...Y_{100}|\theta \overset{\text{iid}}{\sim}$ Bernoulli(θ). Write down the joint distribution of $\Pr(Y_1 = y_1, ...Y_{100} = y_{100}|\theta)$ in a compact form.

(b) Let $X = \sum_i Y_i$. Write down the distribution of X in a compact form.

(c) Suppose that the results of the survey are $X = 66$. Suppose that $\theta \in \{0.0, 0.1, ...0.9, 1.0\}$. Compute $\Pr(X = 66|\theta)$ (the likelihood) for each of these 11 values of θ and plot these probabilities as a function of θ.

(d) Now suppose $\theta \in [0, 1]$. Using the uniform prior density for θ so that $\pi(\theta) = 1$, write down the posterior density $\pi(\theta|X = 66)$.

(a) Given that the $Y_1, Y_2, \cdots, Y_{100}$ are conditional iid on θ,

and they are follow Bernoulli distribution with θ.

$$P(Y_1 = y_1, \cdots, Y_{100} = y_{100}|\theta) = \theta^{\sum_{i=1}^{100} y_i} (1-\theta)^{100-\sum_{i=1}^{100} y_i}$$

(b) Given than $X = \sum_i Y_i$, X follows a Binomial distribution $\sim Bin(X, 100, \theta)$

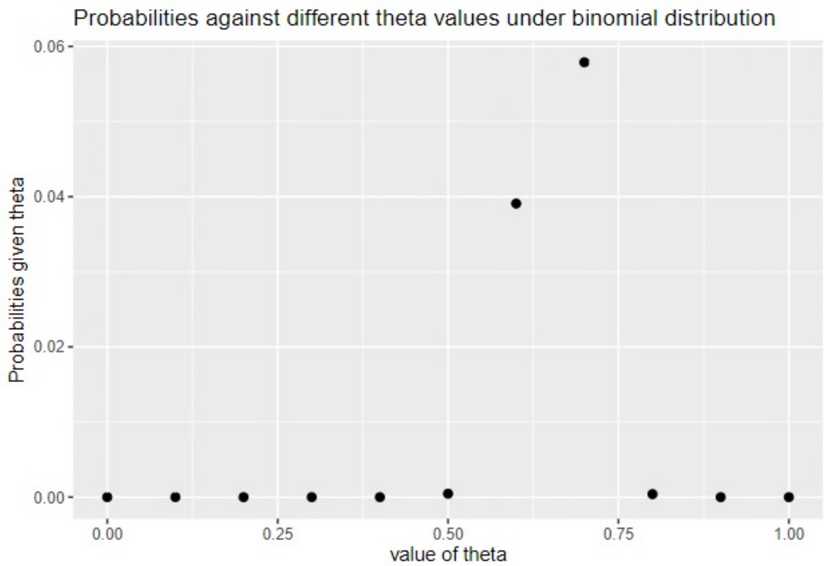
$$P(X = \sum_i Y_i | \theta) = \binom{100}{x} \theta^x (1-\theta)^{100-x}$$

(c) $$P(X = \sum_i Y_i = 66 | \theta) = \binom{100}{66} \theta^{66} (1-\theta)^{34}$$

Question 4

Part C

```
theta <- seq(0, 1, by=0.1)
prob_X <- dbinom(66, size = 100, prob = theta)
prob_likelihood <- tibble(theta, prob_X)
ggplot(prob_likelihood, aes(x=theta, y=prob_X)) +
  geom_point(size = 2) + labs(x="value of theta",
  y="Probabilities given theta",
  title="Probabilities against different theta values under binomial distribution")
```



(d) $P(X=66|\theta) = \binom{100}{66}\theta^{66}(1-\theta)^{34}$ for $\theta \in [0,1]$

$\pi(\theta) = 1$

$$\pi(\theta|X=66) = \frac{\theta^{66}(1-\theta)^{34}}{\int_0^1 \theta^{66}(1-\theta)^{34}\,d\theta}$$

$$= \frac{\theta^{66}(1-\theta)^{34}}{\frac{\Gamma(66+1)\Gamma(34+1)}{\Gamma(66+34+2)}}$$

$$= \frac{\Gamma(102)}{\Gamma(67)\Gamma(35)} \cdot \theta^{66}(1-\theta)^{34}$$

Therefore, the posterior density $\pi(\theta|X=66) \sim Beta(67, 35)$

Since Uniform $(0,1)$ is also equivalent to Beta $(1,1)$ as prior,

we can also update posterior distribution in this way.

Problem 5. (30 points)

Estimate the probability θ of Covid-19 reinfections based on a study in which there were $n = 43$ previously infected individuals and $y = 15$ re-infected individuals within 36 months.

(a) Using a beta(2, 8) prior for θ, write down $\pi(\theta)$, and $\pi(y|\theta)$, and derive $\pi(\theta|y)$.

(b) Find the posterior mean, mode and standard deviation for θ.

(c) Find a 95% credible interval for θ.

(d) Plot $\pi(\theta)$, $\pi(y|\theta)$, and $\pi(\theta|y)$ as functions of θ.

(e) Repeat a-d using a beta(8, 2) prior for θ.

(f) Consider the following prior distribution for θ:

$$\pi(\theta) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}[3\theta(1-\theta)^7 + \theta^7(1-\theta)],$$

which is a 75-25% mixture of a beta(2, 8) and a beta(8, 2) prior distribution.

(i) Given this prior, write down $\pi(\theta)\pi(y|\theta)$ and simplify as much as possible.

(ii) The posterior distribution is a mixture of two distributions. Identify these distributions.

(a) Prior $\theta \sim \text{Beta}(2,8)$

$$\pi(\theta) = \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\theta^1(1-\theta)^7$$

$$\pi(y|\theta) = \binom{43}{15}\theta^{15}(1-\theta)^{28}$$

Since the Prior is a Beta distribution and likelihood function is a binomial the Posterior distribution is also a Beta $\sim (2+15,\ 8+43-15)$

$\sim (17,\ 36)$

$$\pi(\theta|y) = \frac{\Gamma(53)}{\Gamma(17)\Gamma(36)}\theta^{16}(1-\theta)^{35}$$

(b) Posterior Mean:

$$E[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n} = \frac{2+15}{2+8+43} = \frac{17}{53}$$

Posterior mode:

$$\frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{\alpha+y-1}{\alpha+\beta+n-2} = \frac{2+15-1}{2+8+43-2} = \frac{16}{51}$$

Posterior Standard deviation

$$\sqrt{\frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}} = \sqrt{\frac{(2+15)(8+43-15)}{(2+8+43)^2(2+8+43+1)}} = \sqrt{\frac{17\times 36}{53^2\times 54}} = \sqrt{\frac{34}{8427}} = 0.064$$

(c) $P(a<\theta<b) = \int_a^b \pi(\theta|y)\,d\theta = 0.95$

Since $\pi(\theta|y) \sim \text{Beta}(17,36)$

We can use R code `qbeta(c(0.025, 0.975),17,36)`

The 95% credible interval for θ is $(0.203,\ 0.451)$

Question 5

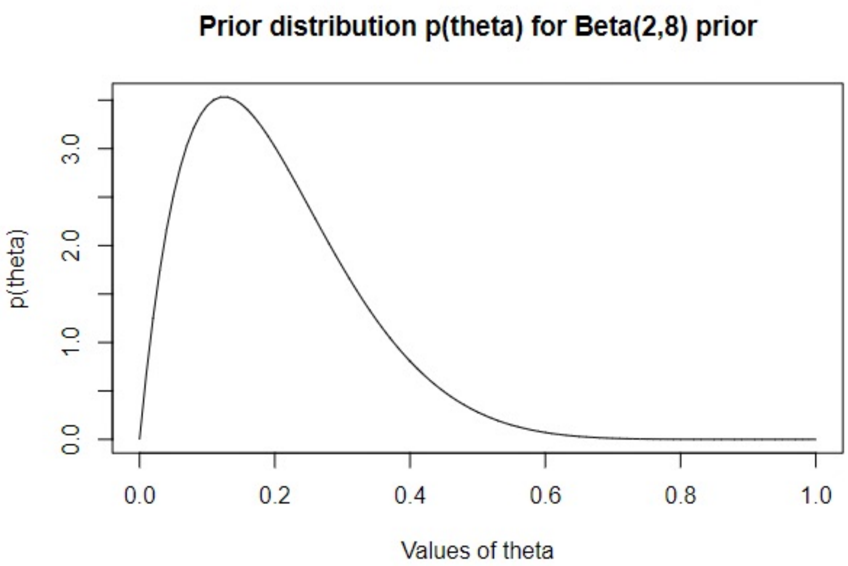
Part C

```
# 95% Credible interval for prior Beta(2,8)
qbeta(c(0.025, 0.975), 17, 36)
```

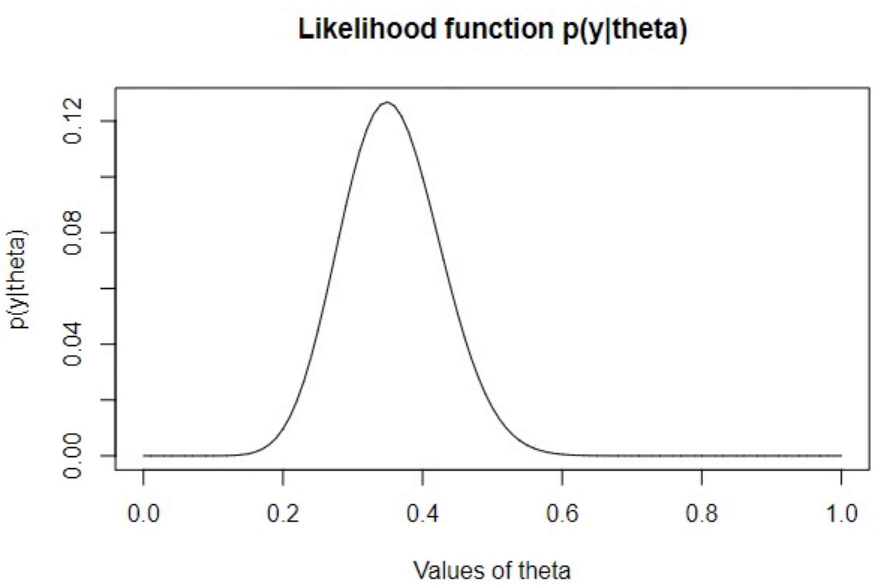
```
## [1] 0.2032978 0.4510240
```

Part D

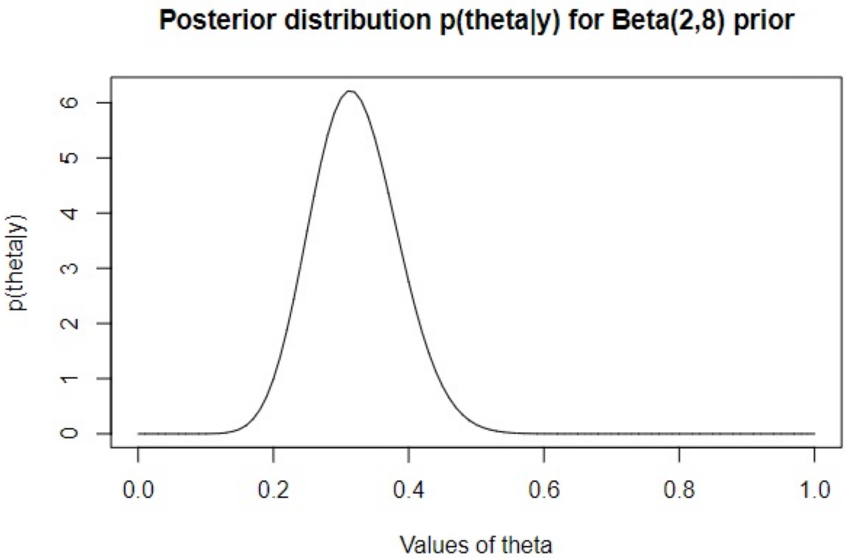
```
# Plot prior, likelihood, and posterior functions for Beta(2,8) prior
y = 15
n = 43
a_1 = 2
b_1 = 8
a_p1 = 17
b_p1 = 36
theta_1 <- seq(0, 1, 0.01)

# Prior distribution p(theta)
plot(theta_1, dbeta(theta_1, shape1 = a_1, shape2 = b_1), type = 'l',
     xlab="Values of theta", ylab="p(theta)",
     main ="Prior distribution p(theta) for Beta(2,8) prior")
```




```
# Likelihood function P(y|theta)
plot(theta_1, dbinom(15, size = 43, prob = theta_1), type = 'l',
     xlab="Values of theta", ylab="p(y|theta)",
     main ="Likelihood function p(y|theta)")
```



```r
# Posterior distribution p(theta/y)
plot(theta_1, dbeta(theta_1, shape1 = a_p1, shape2 = b_p1), type = 'l',
     xlab="Values of theta", ylab="p(theta|y)",
     main ="Posterior distribution p(theta|y) for Beta(2,8) prior")
```



(e) Prior $\theta \sim Beta(8,2)$

$$\pi(\theta) = \frac{\Gamma(10)}{\Gamma(8)\Gamma(2)} \theta^{7}(1-\theta)^{1}$$

$$\pi(y|\theta) = \binom{43}{15} \theta^{15}(1-\theta)^{28}$$

Since the prior is a Beta $(8,2)$ and the likelihood function is a binomial distribution,

the posterior distribution is also a Beta $\sim (8+15, 2+43-15)$

$\sim (23, 30)$

$$\pi(\theta|y) = \frac{\Gamma(53)}{\Gamma(23)\Gamma(30)} \theta^{22}(1-\theta)^{29}$$

Posterior mean:

$$E[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n} = \frac{8+15}{8+2+43} = \frac{23}{53}$$

Posterior mode:

$$\frac{\alpha^*-1}{\alpha^*+\beta^*-2} = \frac{\alpha+y-1}{\alpha+\beta+n-2} = \frac{8+14}{8+2+43-2} = \frac{22}{51}$$

Posterior standard deviation

$$\sqrt{\frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}} = \sqrt{\frac{(8+14)(2+43-14)}{(8+2+43)^2(8+2+43+1)}} = \sqrt{\frac{22\times 31}{53^2\times 54}} = \sqrt{\frac{341}{75843}} = 0.067$$

$$P(a<\theta<b) = \int_a^b \pi(\theta|y)\,d\theta = 0.95$$

Since $\pi(\theta|y) \sim Beta(23, 30)$

we can use R code qbeta(c(0.025, 0.975), 23, 30)

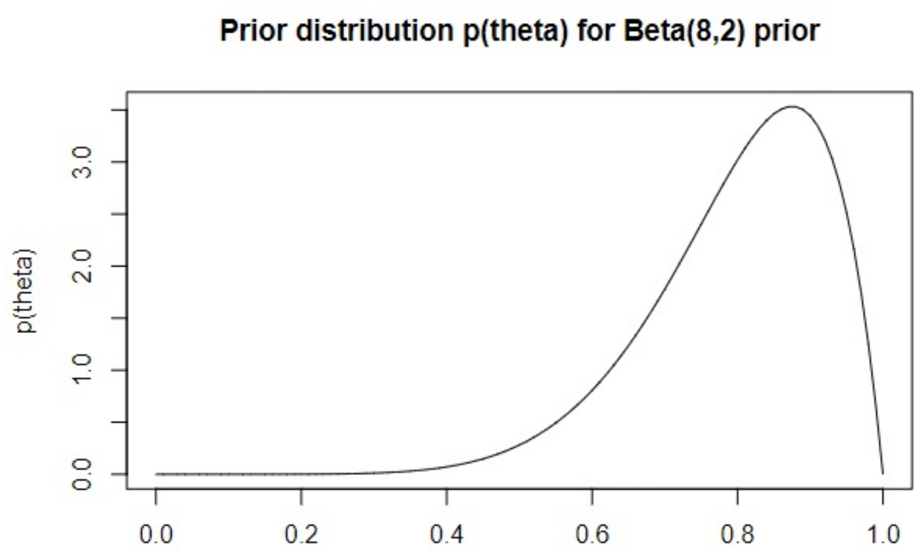
The 95% credible interval for θ is $(0.305, 0.568)$

Part E

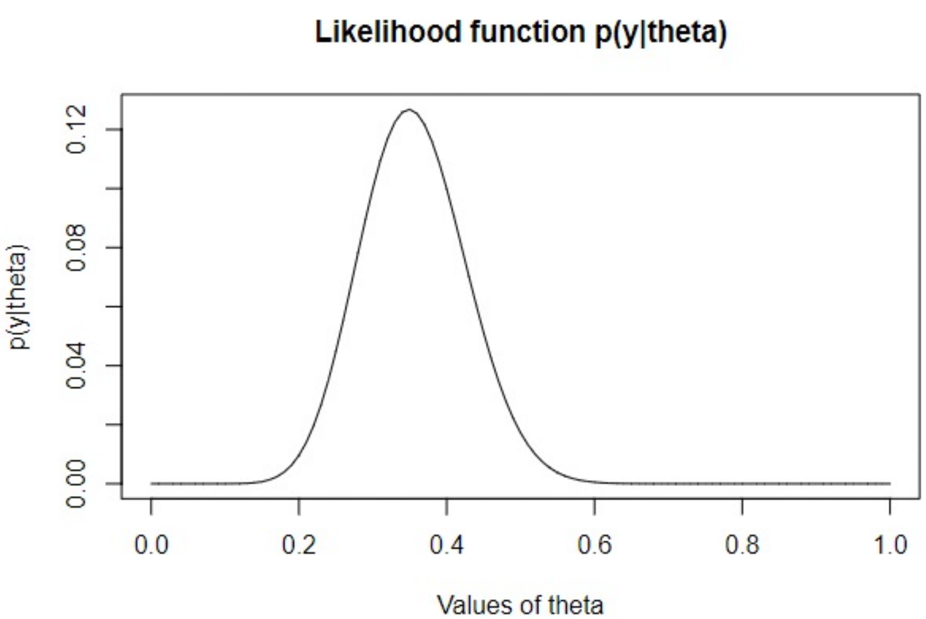
```r
# 95% Credible interval for prior Beta(8,2)
qbeta(c(0.025, 0.975), 23, 30)
```

```
## [1] 0.3046956 0.5679528
```

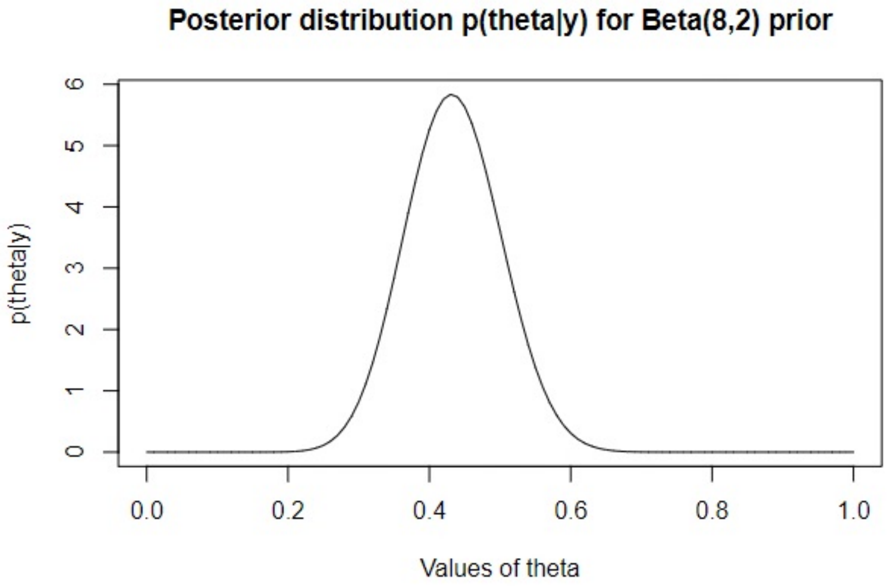
```r
# Plot prior, likelihood, and posterior functions for Beta(8,2) prior
y = 15
n = 43
a_2 = 8
b_2 = 2
a_p2 = 23
b_p2 = 30
theta_1 <- seq(0, 1, 0.01)

# Prior distribution p(theta)
plot(theta_1, dbeta(theta_1, shape1 = a_2, shape2 = b_2), type = 'l',
     xlab="Values of theta", ylab="p(theta)",
     main ="Prior distribution p(theta) for Beta(8,2) prior")
```




```r
# Likelihood function P(y|theta)
plot(theta_1, dbinom(15, size = 43, prob = theta_1), type = 'l',
     xlab="Values of theta", ylab="p(y|theta)",
     main ="Likelihood function p(y|theta)")
```



```r
# Posterior distribution p(theta|y)
plot(theta_1, dbeta(theta_1, shape1 = a_p2, shape2 = b_p2), type = 'l',
     xlab="Values of theta", ylab="p(theta|y)",
     main ="Posterior distribution p(theta|y) for Beta(8,2) prior")
```



f) $\pi(\theta) = \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \left[3\theta(1-\theta)^7 + \theta^7(1-\theta)\right]$

(i) $\pi(\theta) \cdot \pi(y|\theta) \propto \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \left[3\theta(1-\theta)^7 + \theta^7(1-\theta)\right] \cdot \binom{43}{15} \theta^{15}(1-\theta)^{28}$

$$= \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \binom{43}{15} \theta^{15}(1-\theta)^{28} \left[3\theta(1-\theta)^7 + \theta^7(1-\theta)\right]$$

$$= \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \binom{43}{15} \left[3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}\right]$$

$$= \frac{1}{4} \frac{9!}{1!\,7!} \binom{43}{15} \left[3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}\right]$$

$$= 18 \binom{43}{15} \left[3\theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}\right]$$

(ii) The Posterion is a mixture of two Beta distributions. It's Composed by Beta(17, 36) and Beta (23, 30)