R Lab. - Exercise 5

Michele Guadagnini - Mt. 1230663

May 19, 2020

Exercise 1 - Launch of a new journal

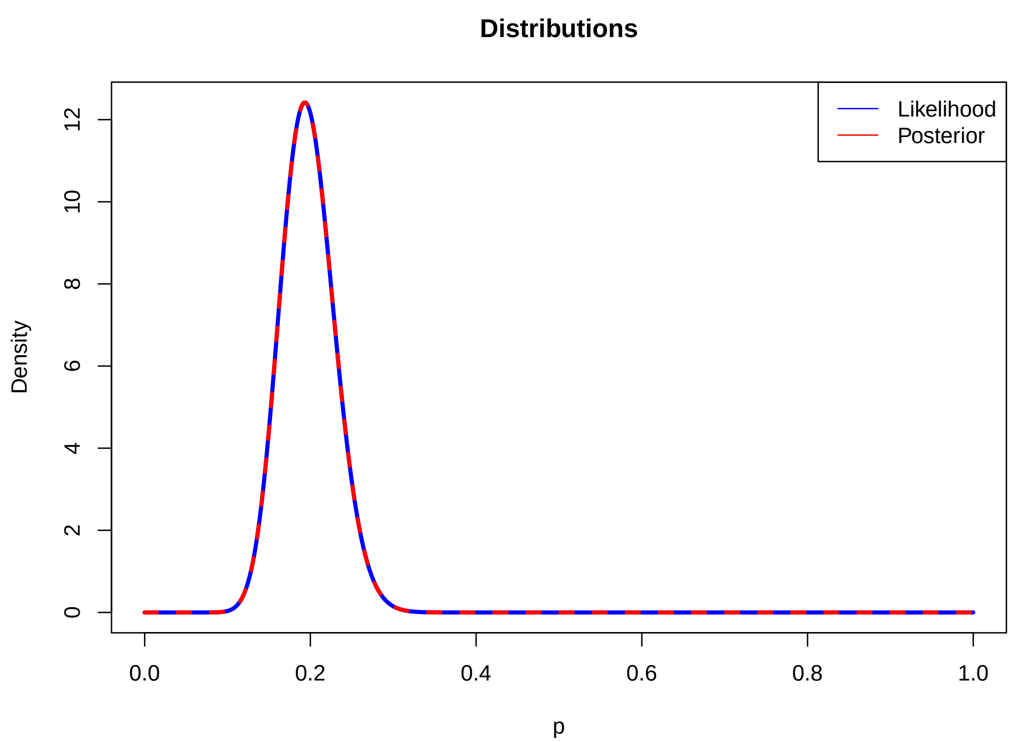
A publishing company has recently launched a new journal. In order to determine how effective it is in reaching its possible audience, a market survey company selects a random sample of people from a possible target audience and interviews them. Out of 150 interviewed people, 29 have read the last issue of the journal.

A) What kind of distribution would you assume for y, the number of people that have seen the last issue of the journal?

B) Assuming a uniform prior, what is the posterior distribution for y?

C) Plot both posterior and likelihood distributions functions.

```
## A ##  Assuming a binomial distribution as likelihood
likelihood <- function(p) { dbinom(29, size=150, prob=p) }
Ilike <- integrate(likelihood, 0, 1)$value
## B ##
prior <- function(p) { dunif(p, 0, 1) }
posterior <- function(p) { likelihood(p)*prior(p) }
Ipost <- integrate(posterior, 0, 1)$value
## C ##
p <- seq(0,1,length=2000)
plot(p,likelihood(p)/Ilike,type='l',col='blue',ylab='Density',lwd=3, main='Distributions')
lines(p, posterior(p)/Ipost, type='l', col='red', lty=2, lwd=3)
legend("topright",legend=c("Likelihood","Posterior"),col=c("blue", "red"),lty=1:1)
```

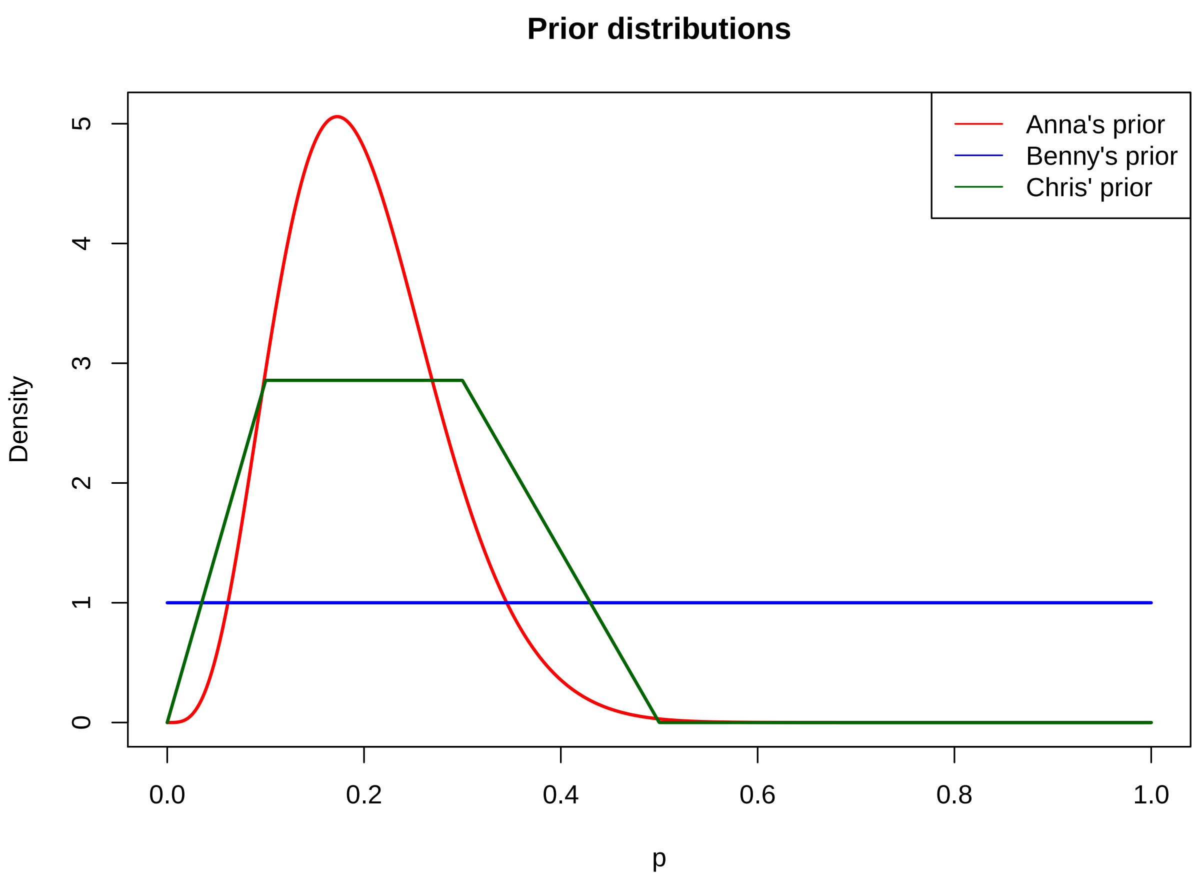


Exercise 2 - New concert hall

Three students want to construct their prior probability about the proportion of residents that support the building of a new concert hall in their small town. Anna thinks that her prior is a beta distribution with mean 0.2 and a standard deviation of 0.08. Benny moved only recently to this new town and therefore he does not have the slightest idea about it. Therefore he decides to use a uniform prior. Chris believes that his prior should have a trapezoidal shape.

A) Draw and compare the three prior distributions.

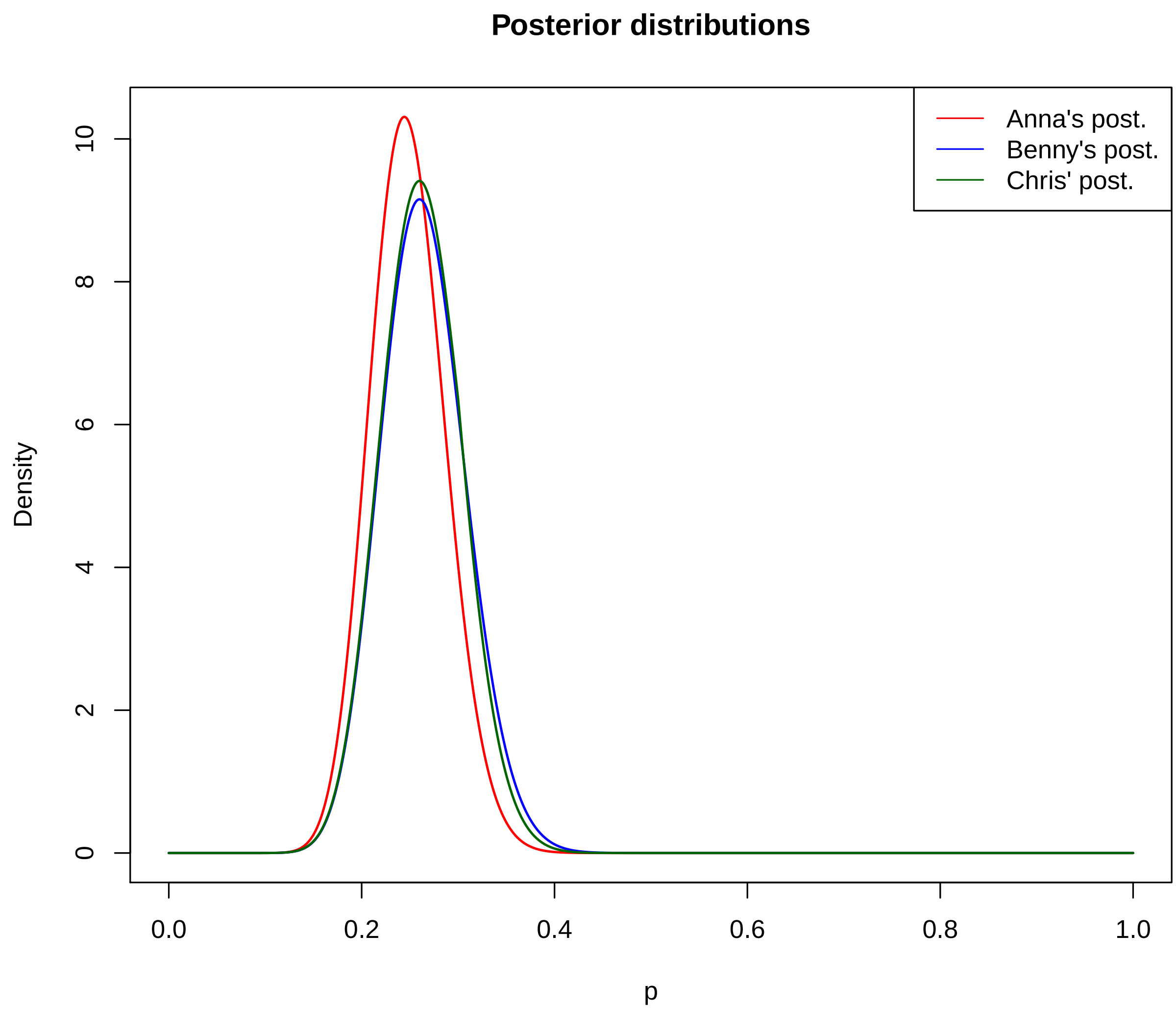
```
mean <- 0.2
var  <- 0.08**2
anna.alpha <- mean*( (mean*(1-mean))/var -1 )
anna.beta  <- (1-mean)*( (mean*(1-mean))/var -1 )
anna.prior <- function(p) { dbeta(p, anna.alpha, anna.beta) }
benny.prior <- function(p) { dunif(p, 0, 1) }
chris.prior <- function(p){ ifelse(p>=0 & p<0.1, 20*p,
                             ifelse(p>=0.1 & p<0.3, 2,
                                    ifelse(p>=0.3 & p<0.5, 5-10*p, 0))) }
p <- seq(0,1,length=2000)
plot(p, anna.prior(p)/integrate(anna.prior,0,1)$value, type='l', col='red', lwd=2,
     ylab='Density',main='Prior distributions')
lines(p, benny.prior(p)/integrate(benny.prior,0,1)$value, type='l', col='blue', lwd=2)
lines(p, chris.prior(p)/integrate(chris.prior,0,1)$value, type='l', col='darkgreen',lwd=2)
legend("topright",legend=c("Anna's prior","Benny's prior","Chris' prior"),
       col=c("red","blue","darkgreen"),lty=1:1)
```



The next day the three students decide to interview a sample of 100 citizens of the small town, asking for their opinion. Out of the interviewed sample, 26 support the building of the new concert hall.

B) Evaluate and draw the three posterior distributions.

```
likelihood <- function(p) { dbinom(26, 100, p) }

anna.post <- function(p) { likelihood(p)*anna.prior(p) }
benny.post <- function(p) { likelihood(p)*benny.prior(p) }
chris.post <- function(p) { likelihood(p)*chris.prior(p) }

Ipost.anna <- integrate(anna.post,0,1)$value
Ipost.benny <- integrate(benny.post,0,1)$value
Ipost.chris <- integrate(chris.post,0,1)$value

plot(p, anna.post(p)/Ipost.anna, type='l', col='red',ylab='Density',
     main='Posterior distributions', lwd=1.5)
lines(p, benny.post(p)/Ipost.benny, col='blue', lwd=1.5)
lines(p, chris.post(p)/Ipost.chris, col='darkgreen', lwd=1.5)
legend("topright",legend=c("Anna's post.","Benny's post.","Chris' post."),
       col=c("red","blue","darkgreen"),lty=1:1)
```



C) Give an estimate of the most probable value and the 95% credibility interval.

```
mode.anna <- p[which.max(anna.post(p))]
mode.benny <- p[which.max(benny.post(p))]
mode.chris <- p[which.max(chris.post(p))]

panna.post  <- Vectorize( function(p) { integrate(anna.post,0,p)$value/Ipost.anna } )
pbenny.post <- Vectorize( function(p) { integrate(benny.post,0,p)$value/Ipost.benny } )
pchris.post <- Vectorize( function(p) { integrate(chris.post,0,p)$value/Ipost.chris } )

low.anna <- p[max(which(panna.post(p)<=0.025))]
upp.anna <- p[min(which(panna.post(p)>=0.975))]

low.benny <- p[max(which(pbenny.post(p)<=0.025))]
upp.benny <- p[min(which(pbenny.post(p)>=0.975))]

low.chris <- p[max(which(pchris.post(p)<=0.025))]
upp.chris <- p[min(which(pchris.post(p)>=0.975))]
```

The most probable value and the 95% credibility intervals are:

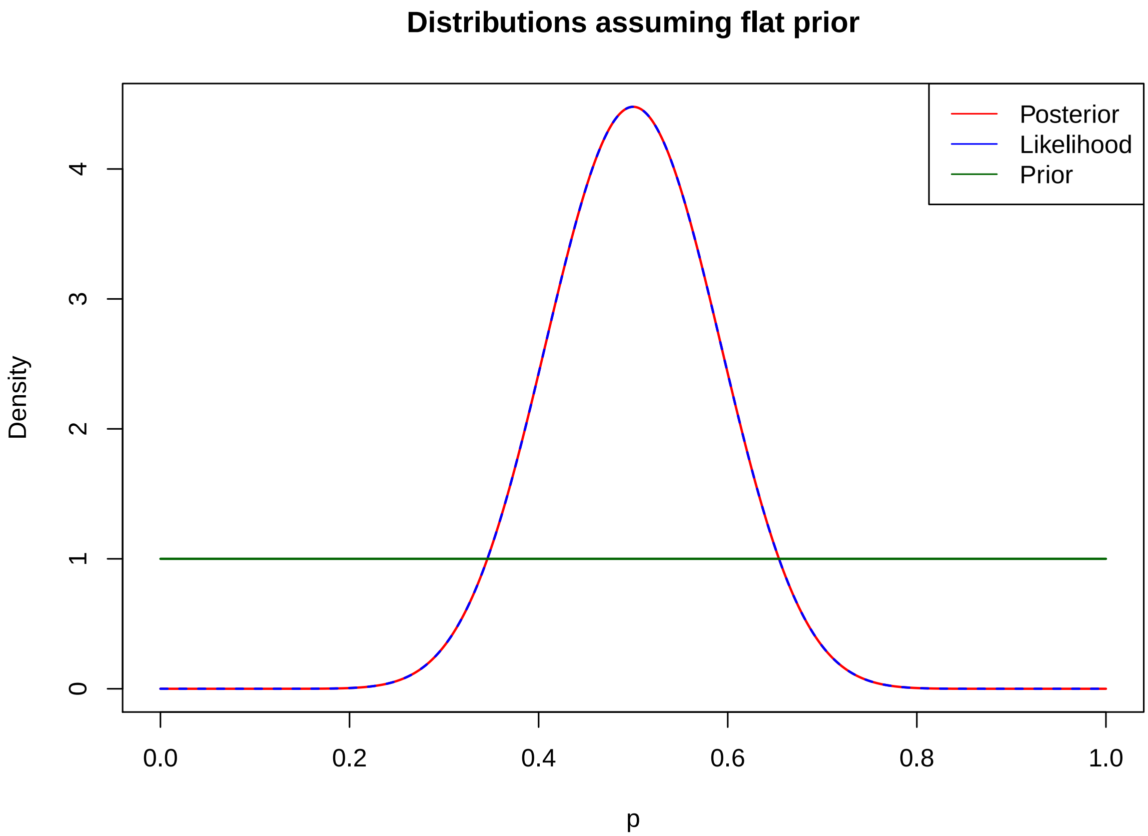
- for Anna : 0.244, [0.177, 0.328]
- for Benny: 0.26, [0.184, 0.354]
- for Chris: 0.26, [0.184, 0.346]

Exercise 3 - Coin flipping

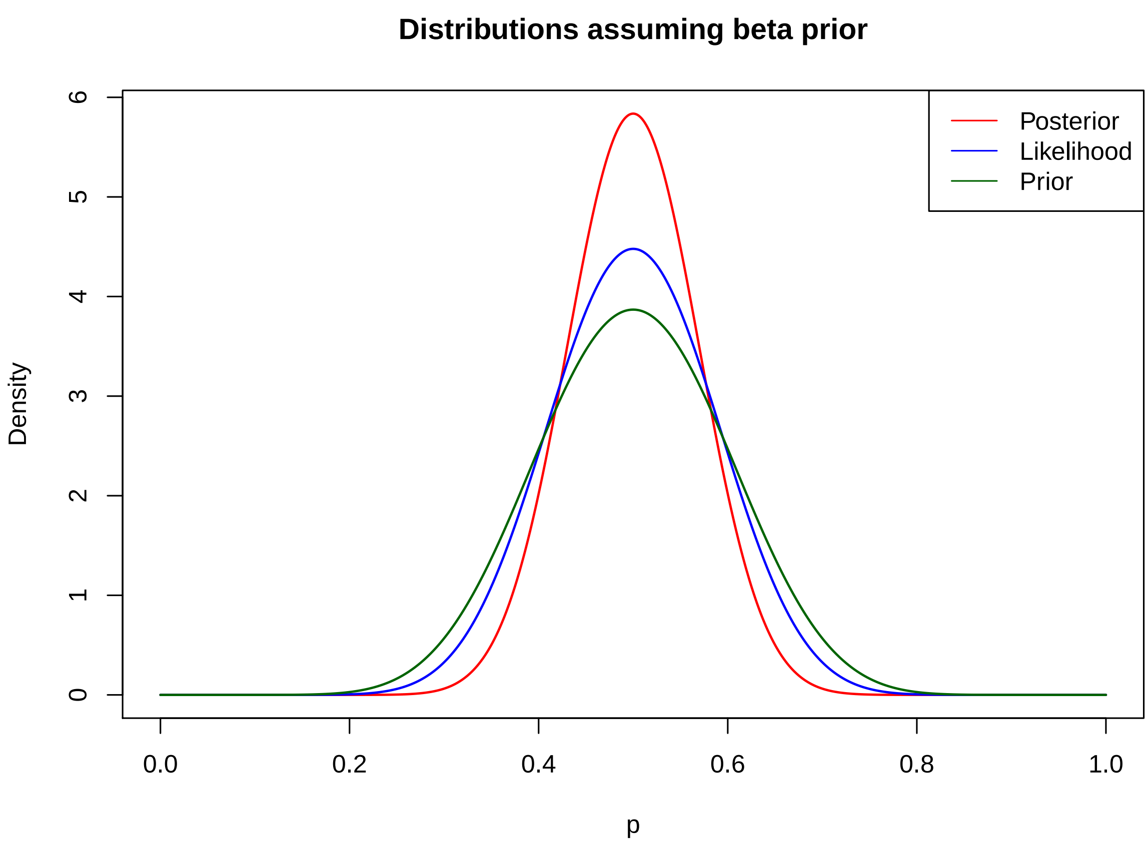
A coin is flipped n = 30 times with the following outcomes:
T, T, T, T, T, H, T, T, H, H, T, T, H, H, H, T, H, T, H, T, H, H, T, H, T, H, T, H, H, H

A) Assuming a flat prior, and a beta prior, plot the likelihood, prior and posterior distributions for the data set.

```
flips <- c('T', 'T', 'T', 'T', 'T', 'H', 'T', 'T', 'H', 'H', 'T', 'T', 'H', 'H', 'H',
           'T',  'H', 'T', 'H', 'T', 'H', 'H', 'T', 'H', 'T', 'H', 'T', 'H', 'H', 'H')
flat.prior <- function(p) { dunif(p, 0, 1) }

mean <- 0.5
var  <- 0.01
alpha <- mean*( (mean*(1-mean))/var -1 )
beta  <- (1-mean)*( (mean*(1-mean))/var -1 )
beta.prior <- function(p) { dbeta(p, alpha, beta) }

r <- length(flips[flips=='H'])
n <- length(flips)
likelihood <- function(p) { dbinom(r, n, p) }

flat.post <- function(p) { likelihood(p)*flat.prior(p) }
beta.post <- function(p) { likelihood(p)*beta.prior(p) }

Ilike <- integrate(likelihood,0,1)$value
Iflat <- integrate(flat.post,0,1)$value
Ibeta <- integrate(beta.post,0,1)$value

par(mfrow=c(2,1))
p <- seq(0,1,length=2000)
plot(p, flat.post(p)/Iflat, col='red', type='l',ylab='Density',
     main='Distributions assuming flat prior',lwd=1.5)
lines(p, likelihood(p)/Ilike, col='blue',lty=2,lwd=1.5)
lines(p, flat.prior(p), col='darkgreen',lwd=1.5)
legend("topright",legend=c("Posterior","Likelihood","Prior"),
       col=c("red","blue","darkgreen"),lty=1:1)

plot(p, beta.post(p)/Ibeta, col='red', type='l',ylab='Density',
     main='Distributions assuming beta prior',lwd=1.5)
lines(p, likelihood(p)/Ilike, col='blue',lwd=1.5)
lines(p, beta.prior(p), col='darkgreen',lwd=1.5)
legend("topright",legend=c("Posterior","Likelihood","Prior"),
       col=c("red","blue","darkgreen"),lty=1:1)
```



Distributions assuming flat prior
Posterior
Likelihood
Prior
Density
p
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0



Distributions assuming beta prior
Posterior
Likelihood
Prior
Density
p
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0

B) Evaluate the most probable value for the coin probability p and, integrating the posterior probability distribution, give an estimate for a 95% credibility interval.

```
mode.flat <- p[which.max(flat.post(p))]
mode.beta <- p[which.max(beta.post(p))]

pflat.post <- Vectorize( function(p) { integrate(flat.post,0,p)$value/Iflat } )
low.flat <- p[max(which(pflat.post(p)<=0.025))]
upp.flat <- p[min(which(pflat.post(p)>=0.975))]

pbeta.post <- Vectorize( function(p) { integrate(beta.post,0,p)$value/Ibeta } )
low.beta <- p[max(which(pbeta.post(p)<=0.025))]
upp.beta <- p[min(which(pbeta.post(p)>=0.975))]
```

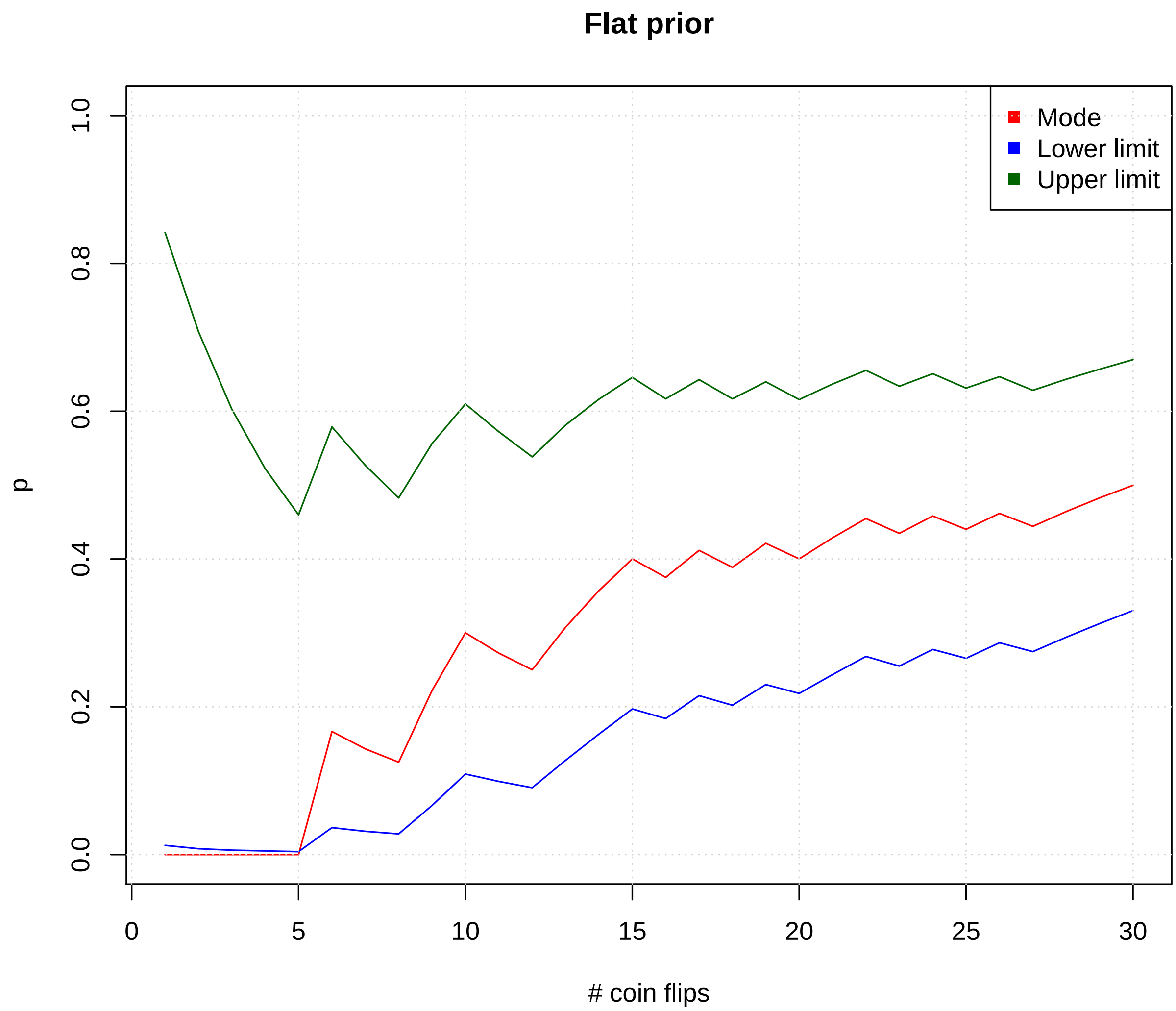
The most probable value for the coin probability and the 95% credibility intervals are:

- with flat prior: 0.5, [0.33, 0.67]
- with beta prior: 0.5, [0.368, 0.632]

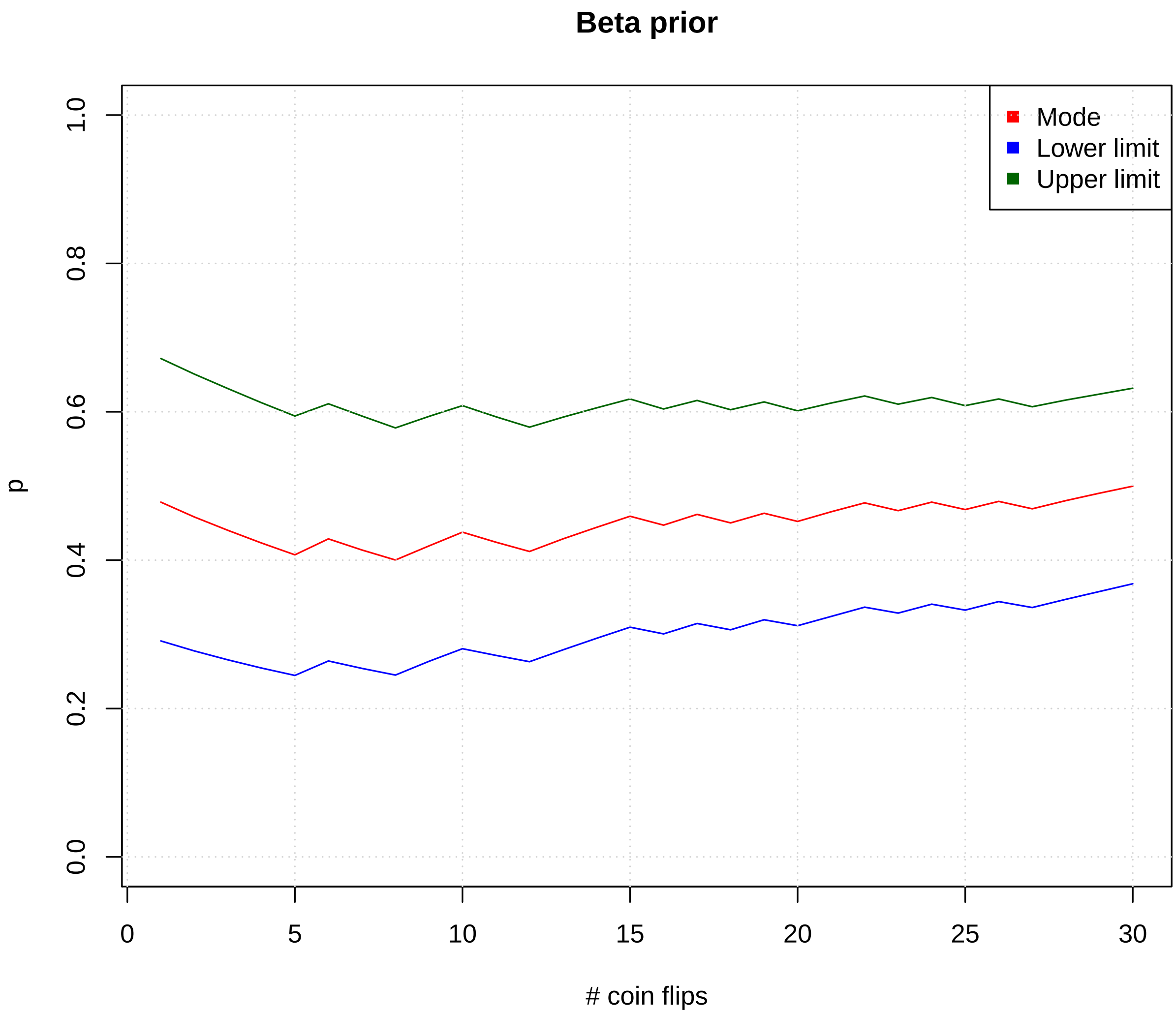
C) Repeat the same analysis assuming a sequential analysis of the data. Show how the most probable value and the credibility interval change as a function of the number of coin tosses (i.e. from 1 to 30).

```
mode.flat <- c()
low.flat <- c()
upp.flat <- c()
mode.beta <- c()
low.beta <- c()
upp.beta <- c()
for (i in 1:30) {
    current_flips <- flips[1:i]
    r <- length(current_flips[current_flips=='H'])
    n <- length(current_flips)

    #a beta distribution with alpha = beta = 1 is equivalent to a uniform distribution.
    #the sequential analysis is obtained updating the parameters (shape1, shape2)
    flat.post <- function(p) { dbeta(p, 1+r, 1+n-r) }

    #a beta dist. is a conjugate prior for a binomial likelihood, so in this case the
    # posterior distribution is still a beta distribution but with updated parameters.
    beta.post <- function(p) { dbeta(p, alpha+r, beta+n-r) }
    Iflat <- integrate(flat.post,0,1)$value
    Ibeta <- integrate(beta.post,0,1)$value

    mode.flat[i] <- p[which.max(flat.post(p))]
    pflat.post <- Vectorize ( function(p) { integrate(flat.post,0,p)$value/Iflat } )
    low.flat[i] <- p[max(which(pflat.post(p)<=0.025))]
    upp.flat[i] <- p[min(which(pflat.post(p)>=0.975))]

    mode.beta[i] <- p[which.max(beta.post(p))]
    pbeta.post <- Vectorize ( function(p) { integrate(beta.post,0,p)$value/Ibeta } )
    low.beta[i] <- p[max(which(pbeta.post(p)<=0.025))]
    upp.beta[i] <- p[min(which(pbeta.post(p)>=0.975))]
}
```

```r
plot(1:30, mode.flat, pch=15, col='red', ylim=c(0,1),xlab='# coin flips',
     ylab='p',main='Flat prior',type='l')
points(1:30, low.flat, pch=15, col='blue',type='l')
points(1:30, upp.flat, pch=15, col='darkgreen',type='l')
legend("topright",legend=c("Mode","Lower limit","Upper limit"),
       col=c("red","blue","darkgreen"),pch=15)
grid()
```




```r
plot(1:30, mode.beta, pch=15, col='red', ylim=c(0,1),xlab='# coin flips',
     ylab='p',main='Beta prior',type='l')
points(1:30, low.beta, pch=15, col='blue',type='l')
points(1:30, upp.beta, pch=15, col='darkgreen',type='l')
legend("topright",legend=c("Mode","Lower limit","Upper limit"),
       col=c("red","blue","darkgreen"),pch=15)
grid()
```



D) Do you get a different result, by analyzing the data sequentially with respect to a one-step analysis (i.e. considering all the data as a whole)?

The most probable value for the coin probability and the 95% credibility intervals at the end of the sequential analysis are:

- with flat prior: 0.5, [0.33, 0.67]
- with beta prior: 0.5, [0.368, 0.632]

These values are equal to the ones obtained before in the one-step analysis.