Alessandro Secchi n. 944668 (code with Manuel Pessina n. 941667)

POLITECNICO
MILANO 1863

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

BAYESIAN LEARNING AND MONTECARLO SIMULATION

HOMEWORK 1

MAY 2020

"A researcher collects data about electrical engineering students and he is interested by estimating the proportion, p, of the number of students that study less than 5 hours per day. Our experimental sample of size 1000 gives us 648 students that study less than 5 hours."

#Part 1 ---

What is the probability distribution of the data? Compute the likelihood and plot it. Note: take everything in percentage. Add a line of the sample proportion to the plot. (Hint: use x = seq(1, 100, 1), size = 100 as parameters in the cumulative distribution). Which continuous probability distribution should be used to describe the prior of this proportion? Specify the function of R and the support.

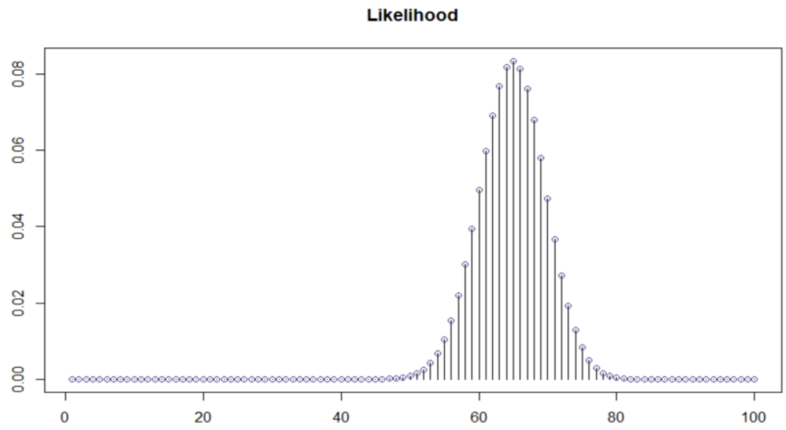
We set the data and before starting we use the command rm(list=ls()) in order to remove all the objects present in the workspace.

```
rm(list=ls())
n<-1000
s<-648
```

Moreover the exercise suggests to take everything in percentage and to add a line of the sample by setting x= seq(1, 100, 1).

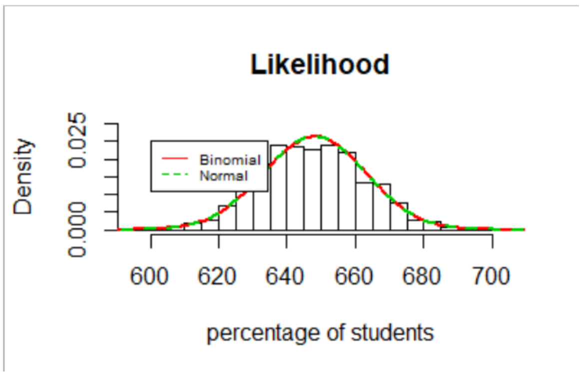
Therefore the likelihood can be written as:

```
plot(dbinom(x, size=100, prob=p),main="Likelihood", xlab = 'percentage of students', col=1, typ="h")
points(x,dbinom(x, size=100, prob=p), col = 'darkslateblue', pch = 1)x= seq(1, 100, 1).
```



Note that if the sample size is large, the sampling distribution of p is approximately a normal (for central limit theorem), but it depends also on p. If p is not close to 0.5 then the sample size should be much larger. Usually a common rule is that if n*p>15 and n(1-p)>15 the normal approximation is acceptable.

In our case with n=1000 we could assume a normal distribution, but we will keep the binomial one, in order to exploit the Beta-Binomial conjugate model.



Note that in this case the formula are: $\mu \sim \mathcal{N}(\mu_0, \tau^2)$ for the prior and for the posterior:

$\mathcal{N}(\mu_n, \sigma_n^2)$ with

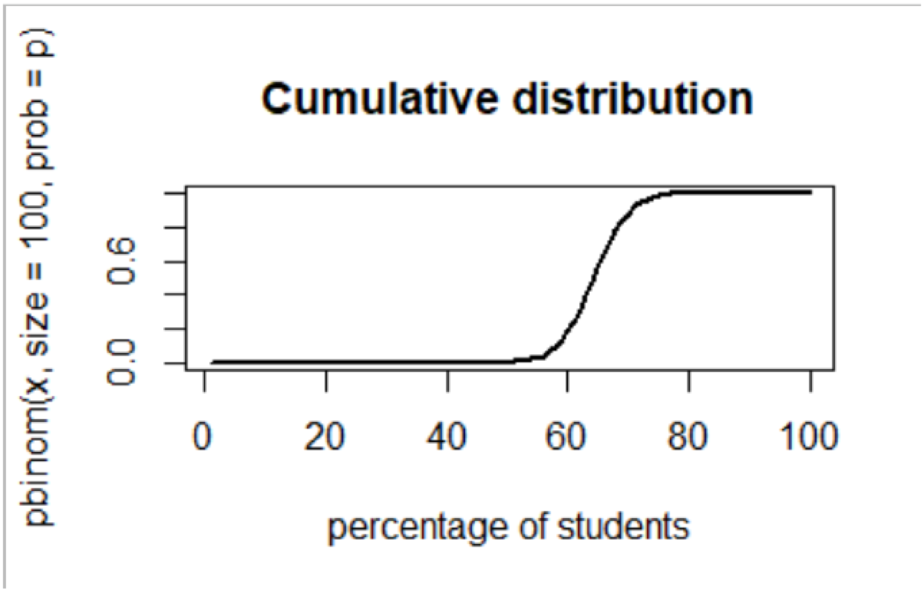
$$\mu_n = \frac{\sigma^2}{\sigma^2 + n\tau^2}\mu_0 + \frac{n\tau^2}{\sigma^2 + n\tau^2}\frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma_n^2 = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

Graphs made by using x=seq(1,1000,1)

```
set.seed(123)
t<-rbinom(x,1000,p)
hist(t,30,prob=T, ylim=c(0,0.030),main="Likelihood", xlab = 'percentage of students')
lines(dbinom(x,1000,p), lwd=2 ,col=2, lty=1)
lines(dnorm(x,mu*n, sigma*n), col=3,add=T, lty=2, lwd=2)
legend(600, 0.025, c("Binomial","Normal"), lty=c(1,2), col=c(2,3), cex=0.6)
```

The cumulative distribution is:

```
plot(pbinom(x,size=100,prob=p),type="l",lwd=2, xlab="percentage of students", main="Cumulative
distribution")
```



This plot shows us the probability that the random variable is less than or equal to a certain amount.

Since we consider the likelihood as a binomial distribution, we use the beta-binomial model. With this assumption we are studying a conjugate model where the posterior will have the same form of the prior.

$$x \sim Bin(n, p) \quad \text{likelihood}$$
$$p \sim \mathcal{B}e(\alpha, \beta) \quad \text{prior}$$

So we wrote our prior as:

```
#prior
curve(dbeta(x, a, b), main="Prior distribution Beta(a,b)", xlab="", ylab=expression(pi))
#support 0,1
```

#Part 2 ---

Another researcher claimed that only 40% of students study less than 5 hours, and this with a variance of 0.2. We want to take this information as a prior for our study. How can we do that? Represent this graphically.

Since we know that the mean and the variance of a Beta distribution are:

$$\boldsymbol{E}[\boldsymbol{x}] = \frac{a}{a+b} = 0.4$$

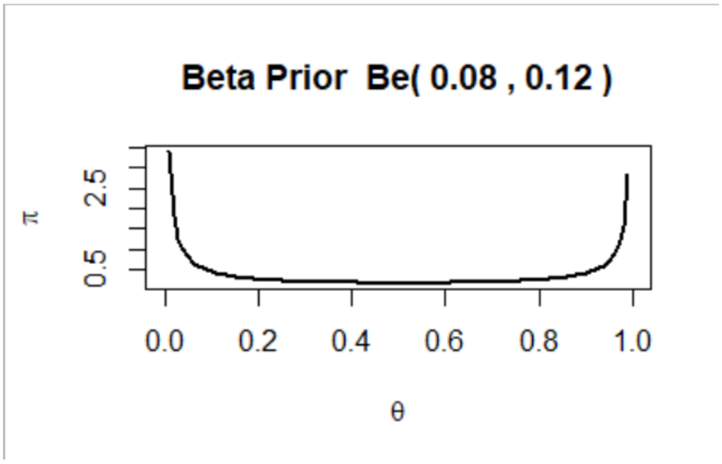
$$\boldsymbol{Var}[\boldsymbol{X}] = \frac{a \cdot b}{(a+b)^2 \cdot (a+b+1)} = 0.2$$

We wrote the code as:

```
mu.new<-0.4
sigma2.new<-0.2
```

Therefore

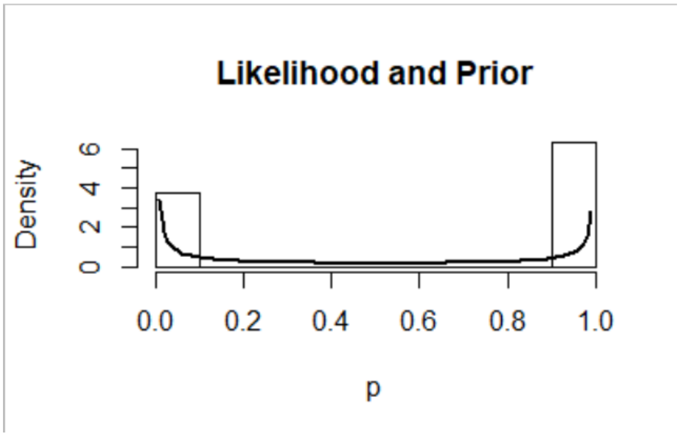
```
a<-12/(125*sigma2.new)-0.4
b<-3/2*a
curve(dbeta(x ,a,b), lty=1,main=paste("Beta Prior"," Be(",a,",",b,")"),
xlab=expression(theta),ylab=expression(pi),lwd=2)
```



More in general we can write these functions:

```
mean.beta <- function(a, b) {a/(a+b)}
variance.beta <- function(a, b) {(a*b)/((a+b)^2*(a+b+1))}
mode.beta <- function(a, b) {if(a>1 & b>1){(a-1)/(a+b-2)} else NA}
```

It is also interesting to notice the following plot by setting size=1 that means we consider only one trial:



With the respective code:

```
set.seed(123)
t=rbinom(x,1,p)
hist(t,10,prob=T,main="Likelihood and Prior", xlab = 'n° students out of 10')
curve(dbeta(x ,a,b),0,1,add=T, lty=1,lwd=2)
```

Here we can see that there is a major influence on the first part of the likelihood (students that study less than 5h) respect to the second one (students who study more than 5h). A result that confirm our p=0.648

#Part 3 ---
Find the posterior distribution of p. Then, plot it together with the prior on the same graph. What do you notice?

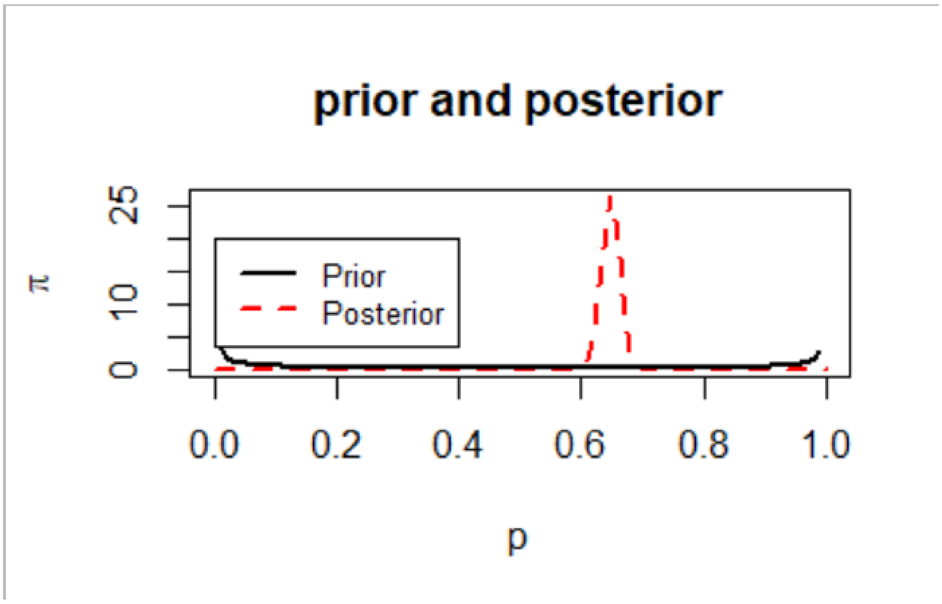
We know that the posterior is:

$$p|x \sim \mathcal{B}e(\alpha + x, \beta + n - x)$$

Therefore:

```
a.post <- a+s
b.post <- b+n-s
curve(dbeta(x ,a.post,b.post),add=F,lty=3,lwd=2)

#plot prior and posterior all toghether
curve(dbeta(x ,a.post,b.post),add=F,lty=2,lwd=2, col=2)
curve(dbeta(x ,a,b), lty=1,add=T,lwd=2, col=1)
legend(0,20,c("Prior", "Posterior"),lty=c(1,2), col=c(1,2), cex=0.8,lwd=2)
title(main='prior and posterior')
```



Since our size is pretty large (n=1000) we can see that the posterior is mainly affected from the likelihood. The prior seems almost uniform, infact having this prior means that I don't give preferences, so the posterior is determined by the data. (which is good since I have n=1000)

Making a little summary we can notice that the posterior mean is 0.648 which corresponds to the MLE estimation (p).

```
mean.prior <- mean.beta(a,b);
var.prior <- variance.beta(a,b);
mode.prior <- mode.beta(a,b)
mean.posterior <- mean.beta(a.post, b.post)
var.post <- variance.beta(a.post,b.post)
mode.post <- mode.beta(a.post,b.post)


(summ <- data.frame(law=c("Prior", "Posterior"),
            mean=c(mean.prior, mean.posterior),
            variance=c(var.prior, var.post),
            mode=c(mode.prior, mode.post),
            mle=c(mle, mle)))
```

And we get:

```
       law      mean     variance      mode   mle
     Prior 0.4000000 0.2000000000        NA 0.648
 Posterior 0.6479504 0.0002278373 0.6482468 0.648
```

Therefore the prior with large variance does not really affect the posterior.

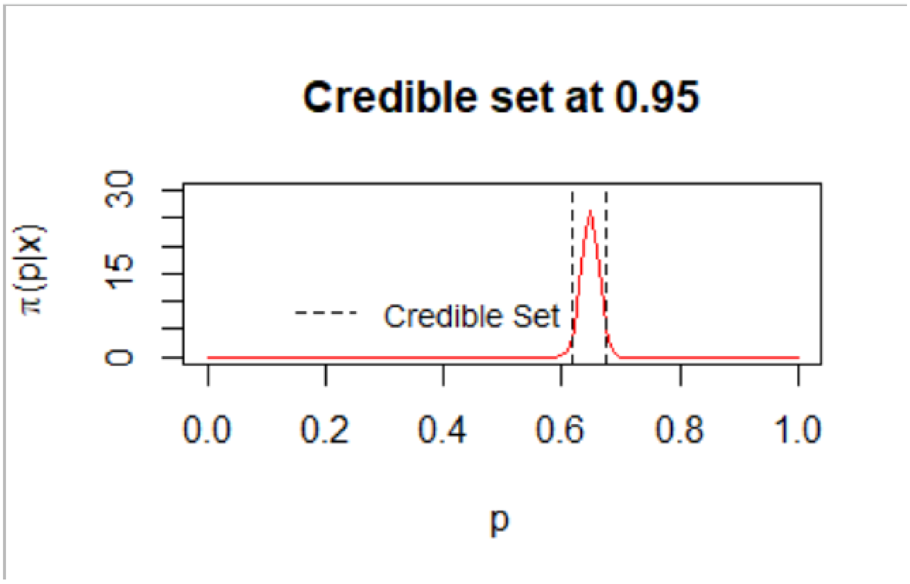
Here we show all the code in only a function:

```
beta.binomial<-function(n,s,a,b){
  mle<-s/n
  a.post <- a+s
  b.post <- b+n-s
  mean.beta<-function(a,b) a/(a+b)
  var.beta<-function(a,b)(a*b)/((a+b)^2*(a+b+1))
  mode.beta<-function(a,b) {if(a>1 & b>1)
  {(a-1)/(a+b-2)} else NA }
  mean.prior<-mean.beta(a,b)
  var.prior<-var.beta(a,b)
  mode.prior<-mode.beta(a,b)
  mean.posterior<-mean.beta(a.post,b.post)
  var.posterior<-var.beta(a.post,b.post)
  mode.posterior<-mode.beta(a.post,b.post)
  summ<-data.frame(law=c("prior","posterior"),
                 mean=c(mean.prior,mean.posterior),
                 variance=c(var.prior,var.posterior),mode=c(mode.prior,mode.posterior),
                 mle=c(mle,mle))
  curve(dbeta(x ,a,b),lty=1, ylim=c(0,25),main=paste("Prior"," Be(",a,",",b,")"),
       xlab=expression(theta),ylab=expression(pi),lwd=2)
   curve(dbeta(x ,a.post,b.post),add=T,lty=3,lwd=2)
  legend(0.1,20,c("Prior","Posterior"),lty=c(1,2,3),cex=0.6,lwd=2)
  summ
}

beta.binomial(n,s,a,b)
```

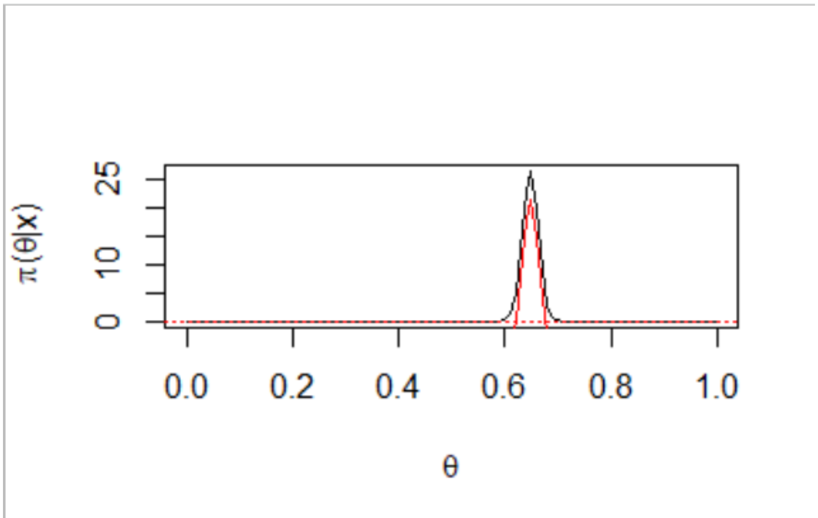
#part4 ---
What are the 95% credible region using HPD and using quantiles? Plot them together with the posterior on the same graph. (Hint: try a sequence from 3 to 6 by 0.1 for h).

```
# credible set (CS) at level 0.95
curve(dbeta(x,a.post,b.post), col=2,ylim=c(0,30),main=paste("Credible set at 0.95"),
     xlab="p",ylab=expression(paste(pi,"(p|x)")),lwd=1)
(CS<-qbeta(c(0.025,0.975),a.post,b.post))
abline(v=CS,lty=2)
legend(0.1,15,"Credible Set",lty=2,bty="n",cex=0.9)
```



While for the HPD we start by setting a threshold of 5 on the densbeta function that we have created.

```
#let's find the zero of the shifted function
densbeta<-function(x,a,b){x^(a-1)*(1-x)^(b-1)/beta(a,b)}
curve(densbeta(x,a.post,b.post),0,1, ylab=expression(paste(pi,"(",theta,"|x)")),xlab=expression(theta))
shift<-function(x,a,b) densbeta(x,a,b)-5
curve(shift(x ,a.post,b.post),lty=1,lwd=1, col=2,add=T)
abline(h=0,lty=3, col=2)
#let's use uniroot function to find the zeros
hpd1<-uniroot(shift,c(0.6,0.65),a.post,b.post)$root
hpd1
hpd2<-uniroot(shift,c(0.65,0.7),a.post,b.post)$root
hpd2
```



Now that we have the two zeros we can integrate to find the area under the curve of beta.

```
integrate(densbeta,lower=hpd1,upper=hpd2,a.post,b.post)
```

We get 0.9320626 with absolute error <9.2e-13 that is less than the 95% therefore we have to take a lower threshold.

Let’s use a for cycle (we used the hint by the exercise using a sequence from 3 to 6 by 0.1 for h).

```
h<-seq(3,6,by=0.1)
results<-matrix(NA,ncol=3,nrow=length(h))
#we set 3 columns cause for every h we find the zero of the shifted function and the value of the
posterior
for(i in 1:length(h)){
  shift<-function(x,a.post,b.post)densbeta(x,a.post,b.post)-h[i]
  hpd1<-uniroot(shift,c(0.6,0.64),a.post,b.post)$root;hpd1
  hpd2<-uniroot(shift,c(0.66,0.7),a.post,b.post)$root;hpd2
  int<-integrate(dbeta,lower=hpd1,upper=hpd2,a.post,b.post)$value
  end<-i
  results[i,]<-c(hpd1,hpd2,int)
  if(int<=0.95) break
}


results[1:end,]
(hpd<-results[end-1,-3])
```

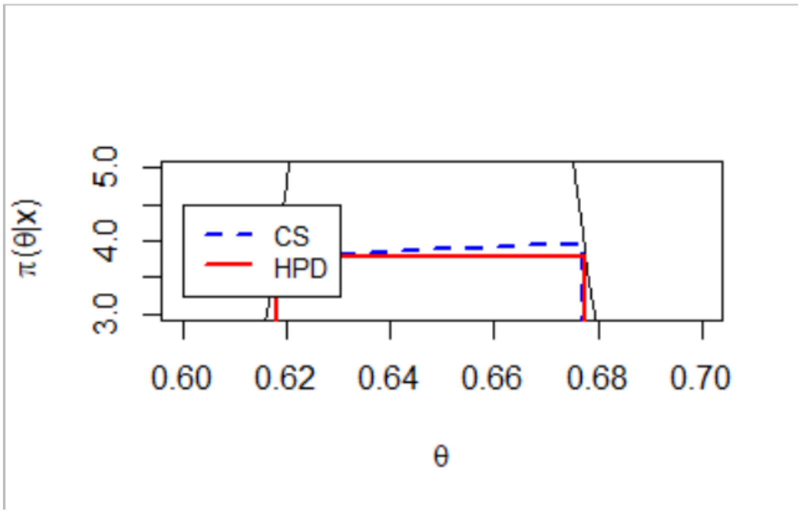
And we get as zeros: 0.6181099 and 0.6775975 with the respective area of 95,12994%

Lastly we plot the two regions together:

```
curve(densbeta(x,a.post,b.post),0,1, ylab=expression(paste(pi,"(",theta,"|x)")),xlab=expression(theta))
#CS
lines( x=c(CS[1],CS[1],CS[2],CS[2]), y=c(0,dbeta(c(CS[1],CS[2]),a.post,b.post),0)  ,col="blue",lwd=2,lty=2)
#HPD
lines( x=c(hpd[1],hpd[1],hpd[2],hpd[2]), y=c(0,dbeta(c(hpd[1],hpd[2]),a.post,b.post),0)  ,col=2,lwd=2)
legend(0.1,20,c("CS","HPD"),col=c("blue",2),cex=0.6,lwd=2,lty=c(2,1))
```

Let’s zoom in order to have a clear plot.

```
curve(densbeta(x,a=a.post,b=b.post),0.6,0.7, ylim=c(3,5),
ylab=expression(paste(pi,"(",theta,"|x)")),xlab=expression(theta))
lines( x=c(CS[1],CS[1],CS[2],CS[2]), y=c(0,dbeta(c(CS[1],CS[2]),a.post,b.post),0)  ,col="blue",lwd=2,lty=2)
lines( x=c(hpd[1],hpd[1],hpd[2],hpd[2]), y=c(0,dbeta(c(hpd[1],hpd[2]),a.post,b.post),0)  ,col=2,lwd=2)
legend(0.6,4.5,c("CS","HPD"),col=c("blue",2),cex=0.8,lwd=2,lty=c(2,1))
```



We see that this Beta distribution is asymmetric and the credible region isn't perfectly a level set.

#part 5 --

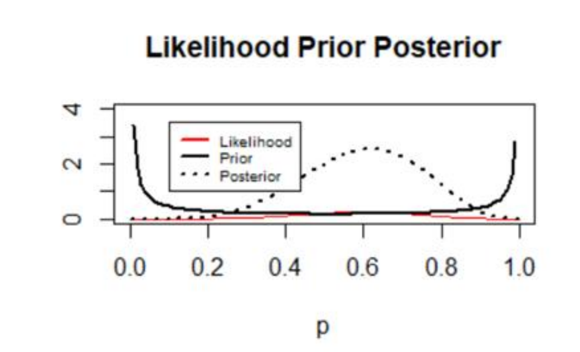
What would happen if we observe a sample of size 10 instead of 1000? And we observe 6 statisticians that study less than 5 hours.

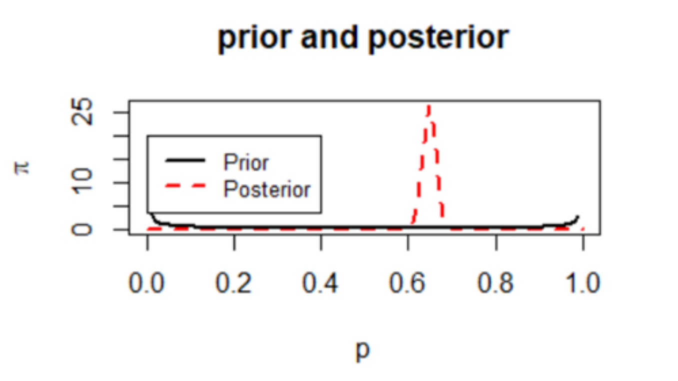
Our new parameters are:

```
n=10
s=6
p=s/n
```

If we keep the same prior as before (a beta distribution with alpha=0.08 and beta=0.12), we get the following plot:

The first is beta.binomial(10,6,0.08,0.12) while the plot at the right is beta.binomial(1000,648,0.08,0.12)





law	mean	variance	mode	mle
prior	0.4000000	0.20000000	NA	0.6
posterior	0.5960784	0.02149723	0.6195122	0.6

law	mean	variance	mode	mle
prior	0.4000000	0.2000000000	NA	0.648
posterior	0.6479504	0.0002278373	0.6482468	0.648

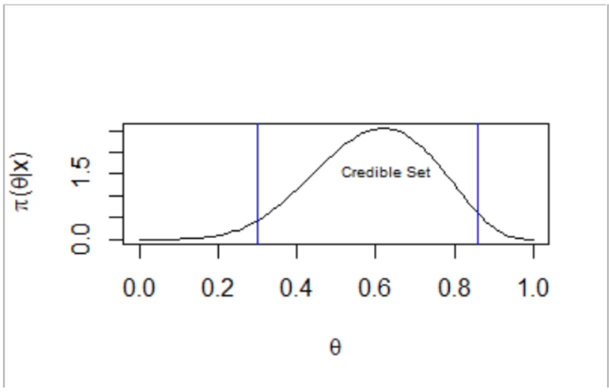
The likelihood has changed and the posterior now is wider than the previous case. Infact the size n=10 is much lower than before and so the posterior variance $\frac{(a+x)(b+n-x)}{(a+b+n)^2(a+b+n+1)}$ is larger.

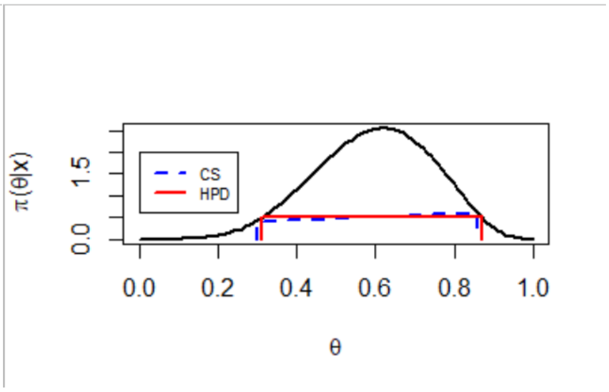
Since we are considering only 10 instead of 1000 students, we can observe a bigger influence on the posterior of the prior.

As we done before we compute the CS and the HPD using the for cycle:

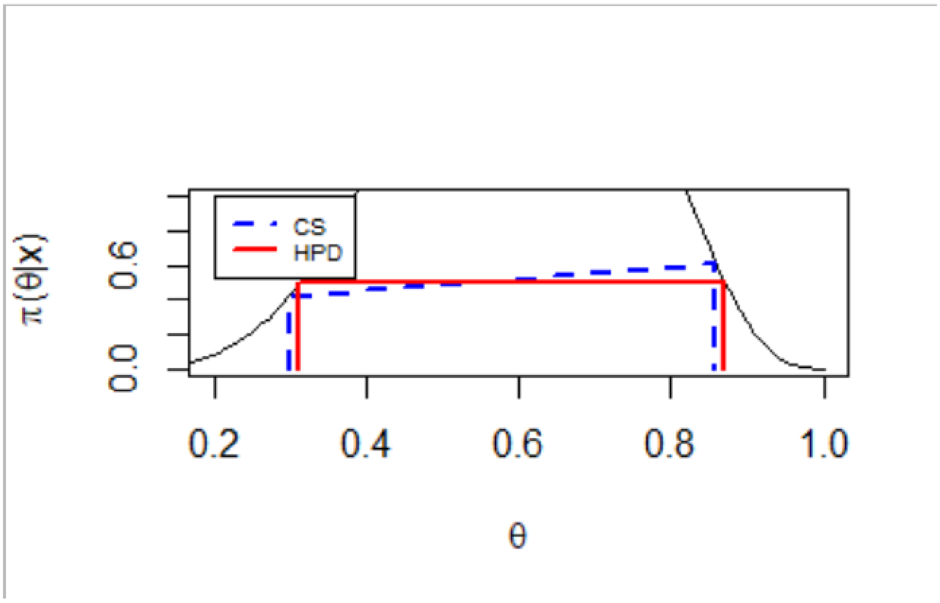
```
densbeta<-function(x,a,b){x^(a-1)*(1-x)^(b-1)/beta(a,b)}
curve(densbeta(x,a=a.post,b=b.post),0,1,ylab=expression(paste(pi,"(",theta,"|x)")),xlab=expression(theta
))

# credible set (CS) at level 0.95
(CS<-qbeta(c(0.025,0.975),a.post,b.post)) # q for quantile
abline(v=CS,lty=2)
legend(0.3,1,"Credible Set",lty=2,bty="n",cex=0.9)
# HPD (Highest Posterior Density)
curve(densbeta(x,a.post,b.post),0,1,ylab=expression(paste(pi,"(",theta,"|x)")),xlab=expression(theta))
shift<-function(x,a,b) densbeta(x,a,b)-0.5
curve(shift(x ,a.post,b.post),add=T,lty=3,lwd=2)
abline(h=0,lty=3)
hpd1<-uniroot(shift,c(0.25,0.35),a.post,b.post)$root
hpd2<-uniroot(shift,c(0.8,0.9),a.post,b.post)$root
integrate(densbeta,lower=hpd1,upper=hpd2,a.post,b.post)
h <- seq(0.4,0.6,by=0.01)
results <- matrix(NA,ncol=3,nrow=length(h))
for(i in 1:length(h)){
  shift<-function(x,a,b)densbeta(x,a,b)-h[i]
  hpd1<-uniroot(shift,c(0.25,0.35),a.post,b.post)$root
  hpd2<-uniroot(shift,c(0.8,0.9),a.post,b.post)$root
  int<-integrate(densbeta,lower=hpd1,upper=hpd2,a.post,b.post)$value
  end<-i
  results[i,]<-c(hpd1,hpd2,int)
  if(int<=0.95) break
}
results[1:end,]
(hpd<-round(results[end-1,-3],6))
curve(densbeta(x,a.post,b.post),0,1,ylab=expression(paste(pi,"(",theta,"|x)")),
    xlab=expression(theta),lwd=2)
lines( x=c(CS[1],CS[1],CS[2],CS[2]), y=c(0,dbeta(c(CS[1],CS[2]),a.post,b.post),0) ,col="blue",lwd=2,lty=2)
lines( x=c(hpd[1],hpd[1],hpd[2],hpd[2]), y=c(0,dbeta(c(hpd[1],hpd[2]),a.post,b.post),0) ,col=2,lwd=2 )
legend(0,2,c("CS","HPD"),col=c("blue",2),cex=0.7,lwd=2,lty=c(2,1))
```





And by zooming we get:



We can observe that the two regions are wider than before since the observations are much lesser than before.

In conclusion, since the posterior variance is $\frac{(a+x)(b+n-x)}{(a+b+n)^2(a+b+n+1)}$, we can state that an higher n leads to a smaller variance of the posterior distribution.