

# lab3

Kerry Hu

## Question 1

Consider the happiness example from the lecture, with 118 out of 129 women indicating they are happy. We are interested in estimating  $\theta$ , which is the (true) proportion of women who are happy. Calculate the MLE estimate  $\hat{\theta}$  and 95% confidence interval.

assume  $Y_i | \theta \sim \text{Bin}(n, \theta)$  where  $y_i$  is the number of women who report to be happy out of the sample of  $n$  women.

$$L(\theta) = \prod_{i=1}^n C_{y_i}^n \theta^{y_i} (1 - \theta)^{(n - y_i)}$$

$$l(\theta) = \log\left(\prod_{i=1}^n C_{y_i}^n \theta^{y_i} (1 - \theta)^{(n - y_i)}\right) = \sum_{i=1}^n \log(C_{y_i}^n) + \sum_{i=1}^n y_i \log(\theta) + \sum_{i=1}^n (n - y_i) \log(1 - \theta)$$

```
library(optimr)
set.seed(1)
data1<-rbinom(200, 129,118/129)
min(data1); max(data1)
```

```
[1] 110
```

```
[1] 124
```

```
Y<-data1
loglike1 <- function( theta, Y) # Assume data are in a vector called Y
{
  n<-129
  log1<- sum(log(factorial(n)/(factorial(n-Y)*factorial(Y))))+sum(Y*log(theta))+sum((n-Y)*
  return(-log1)
} # loglike1 Return minus value of function
```

```
optim(0.5, loglike1, Y=data1)
```

Warning in optim(0.5, loglike1, Y = data1): one-dimensional optimization by Nelder-Mead is un  
use "Brent" or optimize() directly

Warning in log(1 - theta): NaNs produced

Warning in log(1 - theta): NaNs produced

Warning in log(1 - theta): NaNs produced

\$par

[1] 0.9134277

\$value

[1] 492.8948

\$counts

function	gradient
32	NA

\$convergence

[1] 0

\$message

NULL

$\theta_{MLE} = 0.9134277$

Here, assume the MLE of  $\theta$  is asymptotically normal distribution.

```
N<-129
```

```
HH<-optim( 0.5, loglike1, hessian=T,Y=data1 ); HH
```

Warning in optim(0.5, loglike1, hessian = T, Y = data1): one-dimensional optimization by Nel  
use "Brent" or optimize() directly

Warning in log(1 - theta): NaNs produced

```
Warning in log(1 - theta): NaNs produced
```

```
Warning in log(1 - theta): NaNs produced
```

```
$par
```

```
[1] 0.9134277
```

```
$value
```

```
[1] 492.8948
```

```
$counts
```

```
function gradient
```

```
      32      NA
```

```
$convergence
```

```
[1] 0
```

```
$message
```

```
NULL
```

```
$hessian
```

```
      [,1]
```

```
[1,] 326267.1
```

```
thetahat = HH$par[1]
thetahat
```

```
[1] 0.9134277
```

```
SE_thetahat = sqrt(thetahat*(1-thetahat)/N)
SE_thetahat
```

```
[1] 0.02475892
```

```
L_theta = thetahat - 1.96*SE_thetahat; U_theta = thetahat + 1.96*SE_thetahat
```

```
cat("\nEstimated theta = ",round(thetahat,3)," 95 percent CI from ",
+ round(L_theta,3)," to ",round(U_theta,3), "\n")
```

Estimated theta = 0.913 95 percent CI from 0.865 to 0.962

My MLE of  $\theta$  is 0.913 and 95% CI of  $\theta$  is within (0.865,0.962).

## Question 2

Assume a  $\text{Beta}(1,1)$  prior on  $\theta$ . Calculate the posterior mean for  $\hat{\theta}$  and 95% credible interval.

Likelihood is  $Y|\theta \sim \text{Bin}(n, \theta)$  so

$$p(\theta|y) = C_y^n \theta^y (1 - \theta)^{(n-y)}$$

$\theta \sim U(0, 1) = \text{beta}(1, 1)$  so  $p(\theta) = 1$

the posterior distribution

$$p(\theta|y) = \frac{1}{Z} \theta^y (1 - \theta)^{(n-y)}$$

$$Z = \frac{\Gamma(y+1) \Gamma(n-y+1)}{\Gamma(n+2)}$$

The posterior is  $p(\theta|y) \sim \text{Beta}(y+1, n-y+1)$

$$pf = (y+1)/(y+1+n-y+1)$$

```
Y=data1
n=129
#p(theta)=gamma(n+2)/(gamma(Y+1)*gamma(n-Y+1))*theta^Y*(1-theta)^(n-Y)
#a<- gamma(n+2)/(gamma(Y+1)*gamma(n-Y+1))
#p<-gamma(n+2)/(gamma(Y+1)*gamma(n-Y+1))*beta(Y+1,n-Y+1)

pf<- rep(0,200)
pf<-(Y+1)/(n-Y+1+Y+1)
thetahat<-round(mean(pf),3)
thetahat
```

```
[1] 0.907
```

```
CI1<-round(quantile(pf, c(0.025, 0.975)),3)
CI1
```

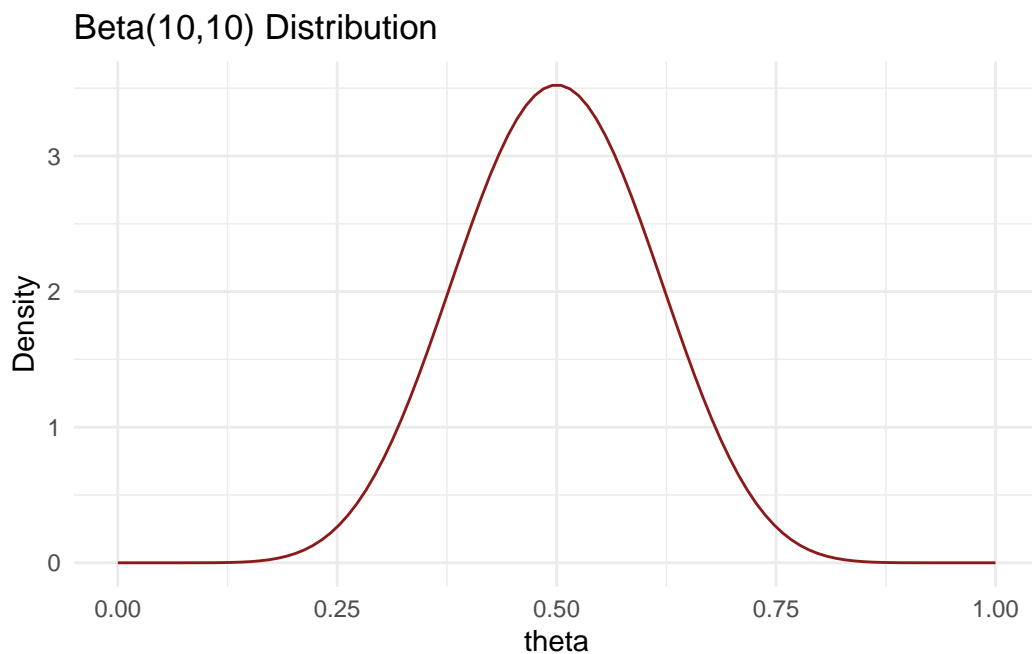
```
2.5% 97.5%
0.862 0.947
```

My the mean of  $\theta$  is 0.907 and 95% CI of  $\theta$  is within (0.862,0.947).

### Question 3

Now assume a  $\text{Beta}(10,10)$  prior on  $\theta$ . What is the interpretation of this prior? Are we assuming we know more, less or the same amount of information as the prior used in Question 2?

```
library(ggplot2)
theta = seq(0, 1, length=100)
df<-dbeta(theta, 10, 10)
df<-as.data.frame(df)
df|> ggplot(aes(x=theta,y=dbeta(theta, 10, 10)))+geom_line(color = "firebrick4")+
  labs(title = "Beta(10,10) Distribution", x="theta",y = "Density")+theme_minimal()
```



```
#In Bayesian language, the prior is your belief about the parameter theta prior to getting
```

The  $\text{Beta}(10,10)$  is a symmetric bell curve to replace a straight line in Question 2. We are assuming we know more amount of information as the prior used in Question 3.

The posterior is

$$p(\theta|y) \sim \text{Beta}(y + 10, n - y + 10)$$

$$ppf = (y + 10) / (y + 10 + n - y + 10)$$

```

Y=data1
n=129
ppf<-(Y+10)/(Y+10+n-Y+10)
thetahat<-round(mean(ppf),3)
thetahat

```

```
[1] 0.858
```

```

CI2<-round(quantile(ppf, c(0.025, 0.975)),3)
CI2

```

```

2.5% 97.5%
0.819 0.893

```

My the mean of  $\theta$  is 0.858 and 95% CI of  $\theta$  is within (0.819,0.893).

#### Question 4

Create a graph in ggplot which illustrates

- The likelihood (easiest option is probably to use `geom_histogram` to plot the histogram of appropriate random variables)
- The priors and posteriors in question 2 and 3 (use `stat_function` to plot these distributions)

Comment on what you observe.

The likelihood function is a binomial distribution.

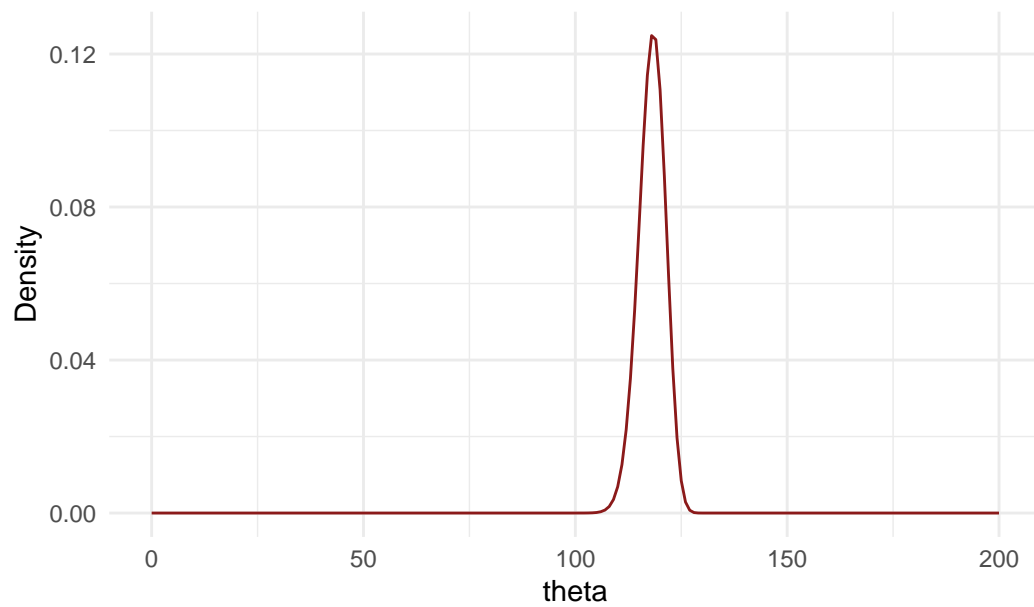
$$p(\theta|y) = C_y^n \theta^y (1 - \theta)^{(n-y)}$$

```

options(scipen=999)
library(ggplot2)
theta = seq(0, 200, by=1)
df<-dbinom(theta,size=200,prob=118/129 )
df<-as.data.frame(df)
df|> ggplot(aes(x=theta,y=dbinom(theta,129 ,118/129 )))+geom_line(color = "firebrick4")+
  labs(title = "The likelihood Binomial Distribution", x="theta",y = "Density")+theme_minimal()

```

## The likelihood Binomial Distribution

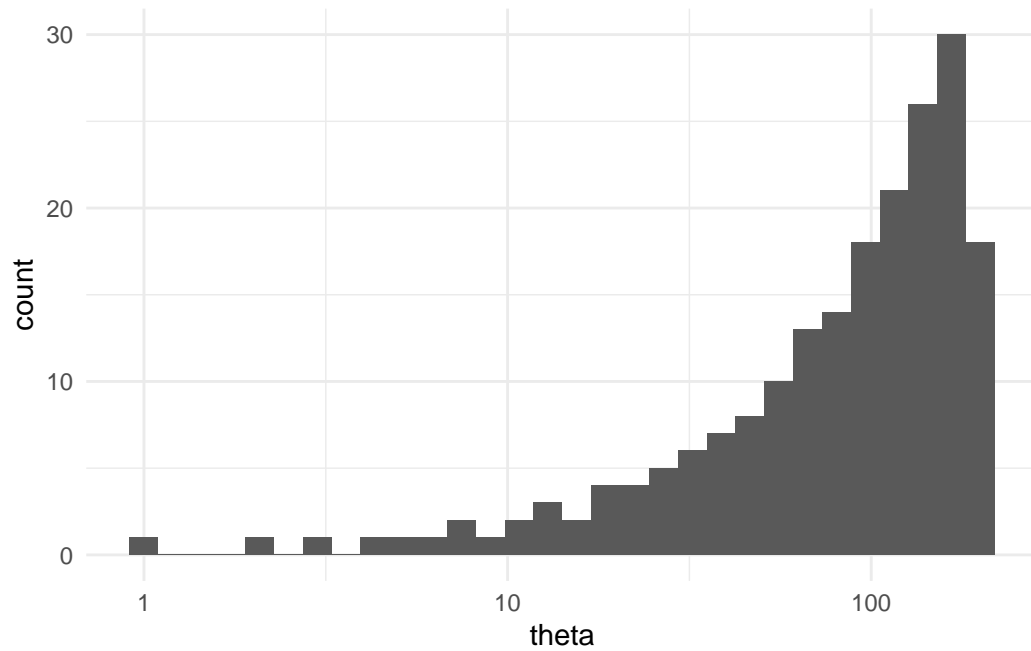


```
df|> ggplot() + geom_histogram(aes(x = theta))+scale_x_log10()+theme_minimal()
```

Warning: Transformation introduced infinite values in continuous x-axis

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning: Removed 1 rows containing non-finite values (`stat\_bin()`).

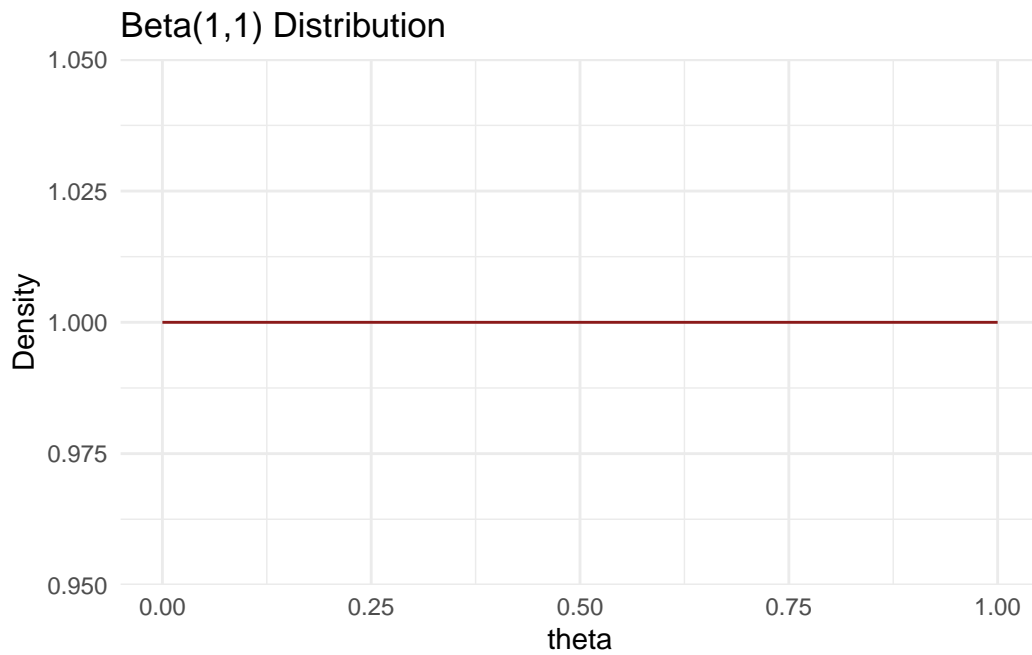


Beta(1,1) prior on  $\theta$  in question 2

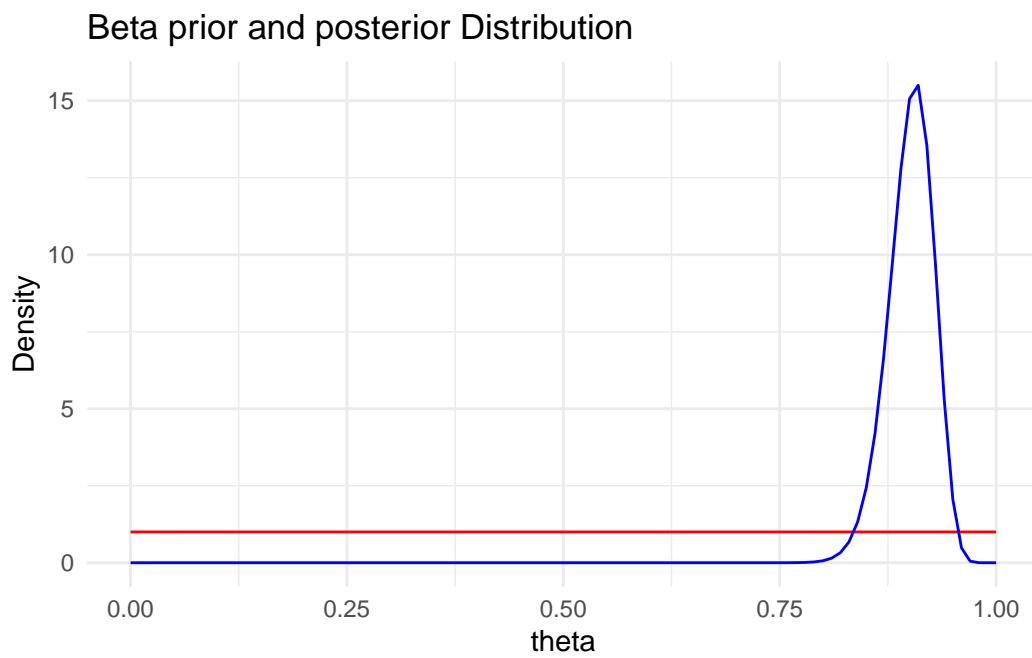
```
library(ggplot2)
Y=as.integer(mean(data1))
n=129
theta = seq(0, 1, length=100)
df1<-dbeta(theta,1,1)
df2<-dbeta(theta,Y+1,n-Y+1)
df<-as.data.frame(cbind(theta,df1,df2))

df|> ggplot(aes(x=theta,y=dbeta(theta, 1, 1)))+geom_line(color = "firebrick4")+
  labs(title = "Beta(1,1) Distribution", x="theta",y = "Density")+theme_minimal()
```



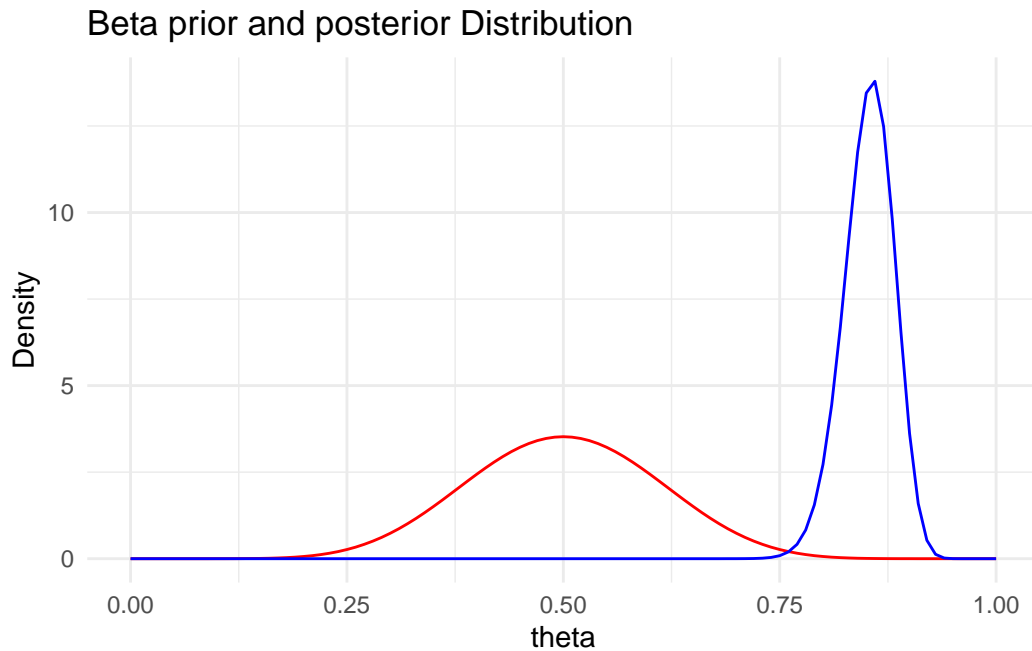


```
ggplot(data.frame(theta= c(0, 1)),aes(theta)) + stat_function(fun = dbeta,args = list(shape
```



Beta(10,10) prior on  $\theta$  and Beta(y+10, n-y+10) posterior on  $\theta$  in question 3

```
library(ggplot2)
Y=as.integer(mean(data1))
n=129
theta = seq(0, 1, length=100)
prior <-stat_function(fun = dbeta,args = list(shape1 = 10, shape2 = 10),color="red")
posterior<- stat_function(fun = dbeta,args = list(shape1 = Y+10, shape2 = n-Y+10),color="blue")
ggplot(data.frame(theta= c(0, 1)),aes(theta),color=prior+posterior) +prior+posterior+labs(
```



In question2 and question3 for the binomial likelihood, a Beta prior results in a Beta posterior distribution: we say that the beta prior is conjugate for the binomial likelihood.

In question2 and question3 no matter their priors are Non-informative prior, Weakly informative, proper and improper, the wave peak values of the final posteriors extremely increased, the location of the bell waves moved to right and values of the theta could close to 0.8-0.9. At the same time the width of the bell waves became relatively narrow.

## Question 5

(No R code required) A study is performed to estimate the effect of a simple training program on basketball free-throw shooting. A random sample of 100 college students is recruited into

the study. Each student first shoots 100 free-throws to establish a baseline success probability. Each student then takes 50 practice shots each day for a month. At the end of that time, each student takes 100 shots for a final measurement. Let  $\theta$  be the average improvement in success probability.  $\theta$  is measured as the final proportion of shots made minus the initial proportion of shots made.

Given two prior distributions for  $\theta$  (explaining each in a sentence):

- A noninformative prior, and
- A subjective/informative prior based on your best knowledge

For simplicity, assume the  $\theta$  are normally distributed with mean  $\mu$  and standard deviation  $\sigma$ .

- 1) Uniform( $0, +\infty$ ) or Beta(1,1) is a Non-informative prior which gives nothing for the data and is improper choice.
- 2) Binomial distribution function is useful informative prior which gives us to compute the proportion for the data.