

# Lab 3: Intro to Bayes

Toufic Ayoub

28/01/23

## 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.

We have that  $Y|\theta \sim \text{Bin}(n, \theta)$ , so the log likelihood is given by

$$\begin{aligned}\ell(\theta; y) &= \log(L(\theta; y)) = \log(f(y|\theta)) \\ &= \log\left(\binom{n}{y}(\theta)^y(1-\theta)^{n-y}\right) \\ &= \log\binom{n}{y} + y\log(\theta) + (n-y)\log(1-\theta)\end{aligned}$$

Differentiating and setting to zero,

$$\begin{aligned}\frac{\partial \ell}{\partial \theta} = 0 &= \frac{y}{\hat{\theta}} + \frac{y-n}{1-\hat{\theta}} \\ &= y(1-\hat{\theta}) + \hat{\theta}(y-n) \\ &= y - y\hat{\theta} + y\hat{\theta} - n\hat{\theta} \\ \hat{\theta} &= \frac{y}{n}\end{aligned}$$

To construct a confidence interval, we need the variance of  $\hat{\theta}$ , so we need to find the Fischer information,

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta^2} &= -\frac{y}{\theta^2} + \frac{y-n}{(1-\theta)^2} \\ \frac{\partial^2 \ell}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} &= -\frac{y}{\hat{\theta}^2} + \frac{y-n}{(1-\hat{\theta})^2}\end{aligned}$$

Thus, we have that

$$Var(\hat{\theta}) = - \left( E \left[ -\frac{y}{\hat{\theta}^2} + \frac{y-n}{(1-\hat{\theta})^2} \right] \Big|_{\theta=\hat{\theta}} \right)^{-1} = - \left( E \left[ -\frac{\theta}{\hat{\theta}^2} + \frac{\theta-n}{(1-\hat{\theta})^2} \right] \Big|_{\theta=\hat{\theta}} \right)^{-1} = \left( \frac{n}{\hat{\theta}(1-\hat{\theta})} \right)^{-1}$$

Given that  $y = 118$  and  $n = 129$ , we can find the variance,

```
n = 129
thetaHat = 118/129
var = (n / (thetaHat*(1-thetaHat)))^-1
var
```

```
[1] 0.0006046521
```

A 95% confidence interval is then constructed as follows,

```
thetaHat + c(-1, 1)*1.96*sqrt(var)
```

```
[1] 0.8665329 0.9629244
```

## Question 2

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

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &\propto p(y|\theta)p(\theta) \\ &\propto \binom{n}{y}(\theta)^y(1-\theta)^{n-y} \\ &\propto \binom{n}{y}(\theta)^{y+1-1}(1-\theta)^{n-y+1-1} \end{aligned}$$

We see that the posterior distribution is  $Beta(y+1, n-y+1)$ . The mean for this distribution is given by  $\frac{\alpha}{\alpha+\beta} = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2} = \frac{118+1}{129+2} \approx 0.908$ .

Our credible interval is then simply  $[\theta_{0.025}, \theta_{0.975}]$ , where these are the posterior quantiles.

```
y=118
n=129
```

```
c(qbeta(0.025,y+1,n-y+1), qbeta(0.975,y+1,n-y+1))
```

```
[1] 0.8536434 0.9513891
```

### Question 3

Now assume a  $Beta(10,10)$  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?

Here the interpretation is the number of successes and failures are balanced out, since  $\alpha$  and  $\beta$  are equivalent. The density shifts cancel one another and we have symmetric distribution. Now even though the two priors are symmetric, we are assuming to have more information here. Recall that the  $Beta(1,1)$  distribution *IS* the uniform distribution, so we haven't really *observed* anything. The uniform prior is *uninformative*. With  $Beta(10,10)$ , the distribution is no longer uniform, so this is only reasonable provided we have some indication it is true, i.e. we have more information.

### Question 4

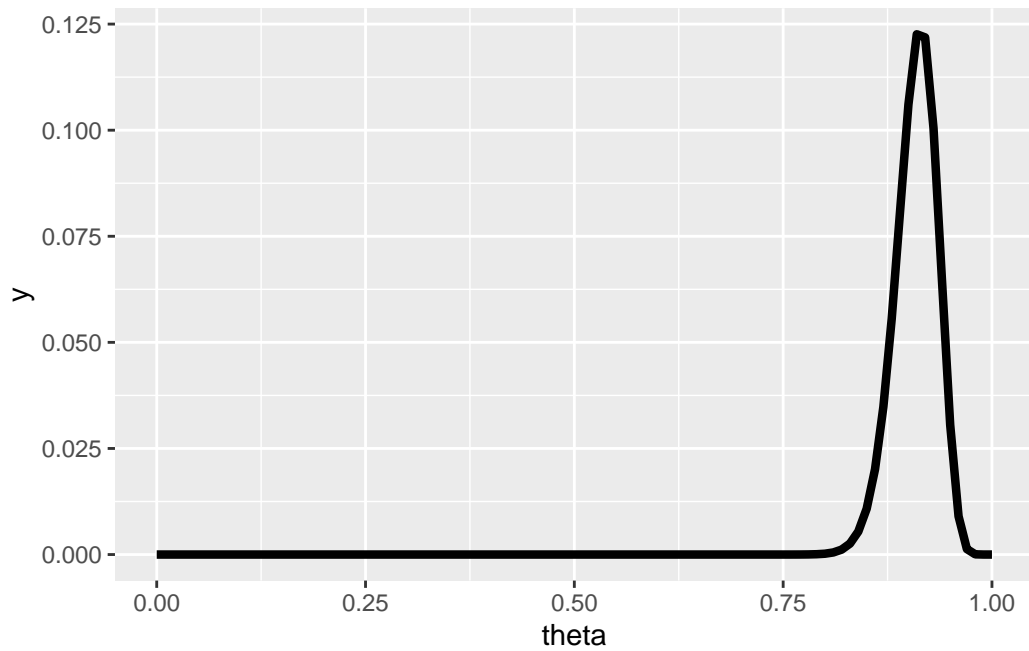
Create a graph in ggplot which illustrates

- 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.

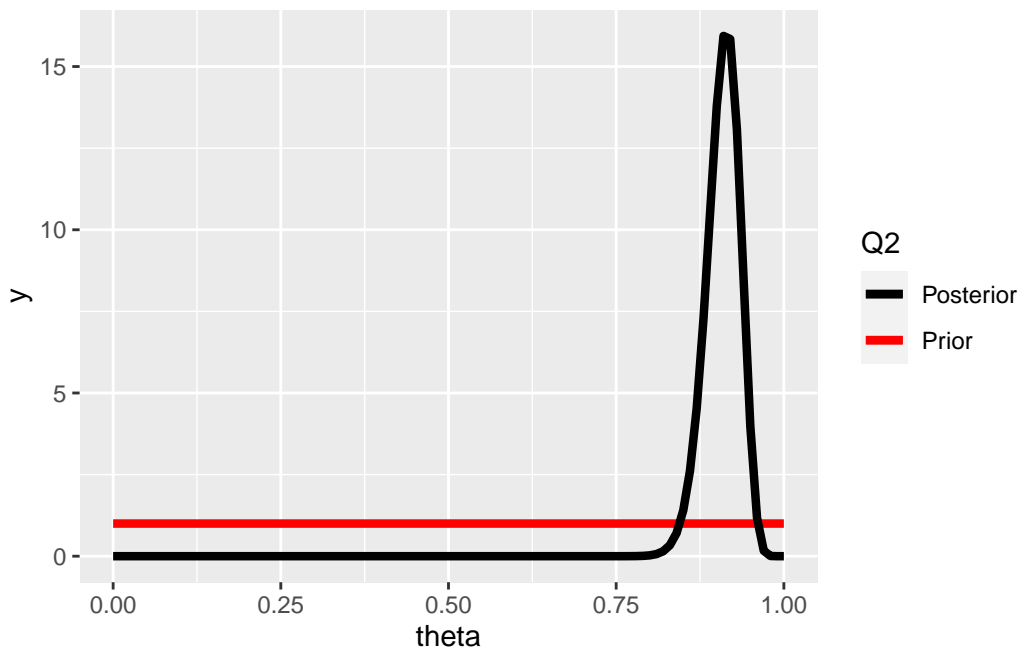
```
library(ggplot2)
theta = seq(0,1,by=0.01)
df<-data.frame(theta)

likelihood = function(theta){
  n = 129
  y = 118
  choose(n,y)*(theta^y)*(1-theta)^(n-y)
}
# Likelihood
ggplot(df, aes(x=theta)) + stat_function(fun=likelihood, lwd=1.5)
```

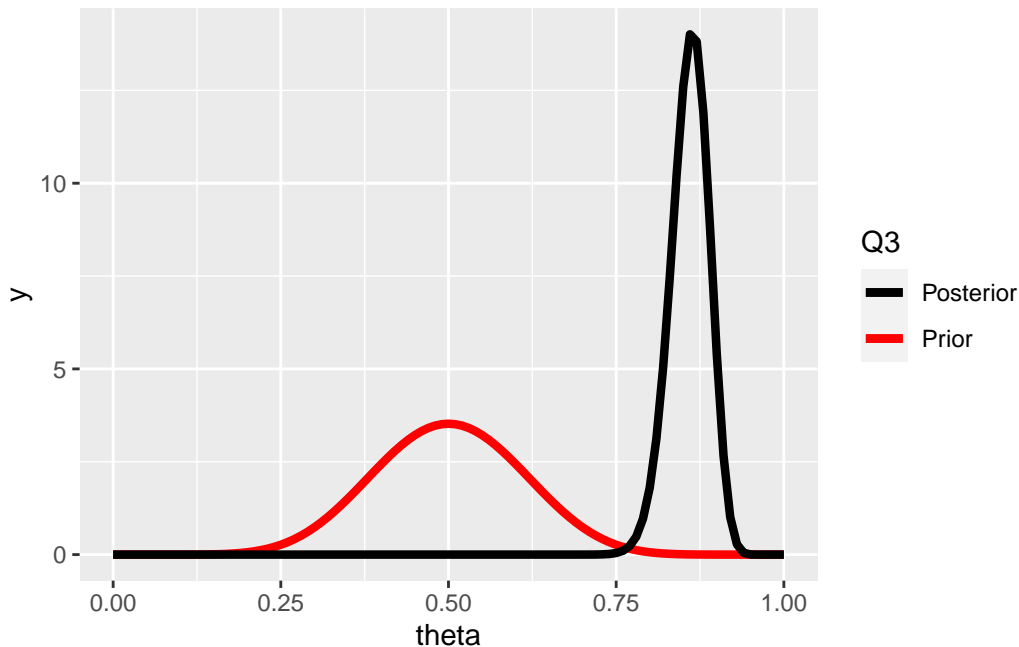


#Q2

```
ggplot(df, aes(x=theta)) + stat_function(fun=dbeta, args = c(1,1), lwd=1.5,
                                         aes(colour = "Prior")) +
  stat_function(fun=dbeta, args = c(y+1,n-y+1), lwd=1.5, aes(colour = "Posterior"))+scale_y
```



```
#Q3
ggplot(df, aes(x=theta)) + stat_function(fun=dbeta, args = c(10,10), lwd=1.5,
                                         aes(colour = "Prior")) +
  stat_function(fun=dbeta, args = c(y+10,n-y+10), lwd=1.5, aes(colour = "Posterior"))+scale_y_continuous(limits=c(0,15))
```



Comparing the prior-posterior pairs, we see that when we looking at the  $Beta(1,1)$  prior, it's posterior density is shifted rightward compared with the posterior density from the  $Beta(10,10)$  prior. From this example then, when we assume more information, the density of  $\theta$  is closer to the left, compared to when we are uninformed. That is, if we assume more information, females are less happy.

## 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:

A uniform distribution say  $Unif(-1, 1)$ . Here we make zero assumptions of any baseline ability students have or the fact that they are training. So all possibilities are equally likely.

- A subjective/informative prior based on your best knowledge:

A beta distribution, say  $Beta(20, 80)$ . Given that the students are practicing every day for a month, and that for a random student, they will likely improve compared to their own baseline regardless of their athleticism; and so we expect  $\theta$  to be positive and be between zero and 0.5 (since they cannot drastically improve that much that quickly).