

# Week 3 Lab

Yeonjoon Choi

2023-01-27

## 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.

Let  $X_i$  be a Bernoulli variable with probability of success  $\theta$ . So we're assuming that each 129 woman have the same chance to be happy, with probability  $\theta$ .

Then, the likelihood is  $L(\theta|x_1, \dots, x_{129}) = \prod_{i=1}^{129} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{129} x_i} (1-\theta)^{129 - \sum_{i=1}^{129} x_i}$ , where  $x_i = 1$  if the  $i$ -th woman is happy, and 0 otherwise. We know that  $\sum_{i=1}^{129} x_i = 118$ , so we may write the likelihood as

$$L(\theta) = \theta^{118} (1-\theta)^{11}$$

The log-likelihood is  $118 \log \theta + 11 \log(1-\theta)$ , and differentiating, we obtain  $\frac{118}{\theta} - \frac{11}{1-\theta}$ . Setting  $\frac{118}{\theta} - \frac{11}{1-\theta} = 0$ , we get  $\theta = \frac{118}{129} \approx 0.9147$ .

The second derivative is  $-\frac{118}{\theta^2} - \frac{11}{(1-\theta)^2} < 0$  at  $\theta = \frac{118}{129}$ . Hence, the maximum likelihood estimate is  $\hat{\theta} = \frac{118}{129}$ .

To construct the confidence interval, we will use the asymptomatic normality of maximum likelihood estimator.

The fisher information for one sample is  $-E(\frac{\partial^2}{\partial \theta^2} \log f(X; \theta)) = E(\frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2}) = \frac{1}{\theta} + \frac{1}{(1-\theta)} = \frac{1}{\theta(1-\theta)}$ .

So the 95% confidence interval will be given by  $\hat{\theta} \pm 1.96 \frac{1}{\sqrt{129 \frac{1}{\hat{\theta}(1-\hat{\theta})}}}$ .

The confidence interval is then (0.8665, 0.9629).

## Question 2

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

Beta(1,1) has probability density function of  $f(x) = 1$  for  $x \in (0, 1)$ .

Hence, the posterior distribution of  $\theta$  given  $x_1, \dots, x_{129}$  is proportional to  $\theta^{\sum_{i=1}^{129} x_i} (1-\theta)^{129 - \sum_{i=1}^{129} x_i}$ , which means it will follow beta distribution with  $\alpha = \sum_{i=1}^{129} x_i + 1$  and  $\beta = 129 + 1 - \sum_{i=1}^{129} x_i$ .

The mean of beta distribution is  $\frac{\alpha}{\alpha+\beta}$ . To see this, note that  $\int_0^1 x^\alpha (1-x)^{\beta-1}$  is proportional to the probability density function of  $Beta(\alpha+1, \beta)$ . And hence,  $\int_0^1 x^\alpha (1-x)^{\beta-1} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$ . Hence, the mean is  $\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$ , so the posterior mean is  $\frac{119}{119+12} \approx 0.9083$ .

So with our data, the posterior distribution is  $Beta(119, 12)$ . From  $R$ , we can find the 95 % credible interval

```
lb = qbeta(0.025, 119, 12)
ub = qbeta(0.975, 119, 12)

#the credible interval
print (c(lb, ub))
```

```
## [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?

We may interpret the prior in the sense that in a previous experiment, we had 9 successes (a woman says she is happy) and 9 failures (a woman says she is unhappy). We're assuming that we know more. In Beta(1,1) prior, we're saying we have no prior information, since we have 0 success and 0 failure. But in Beta(10,10) distribution, we have prior information of 9 success and 9 failure, so we have more certainty in  $\theta$  being around 0.5. We can see this from the graph of probability density function of Beta(1,1) and Beta(10,10).

```
x = seq(from = 0, to =1, by = 0.001)

y_1 = dbeta(x, 1,1)
y_2 = dbeta(x,10,10)

require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Warning: package 'tibble' was built under R version 4.2.2
```

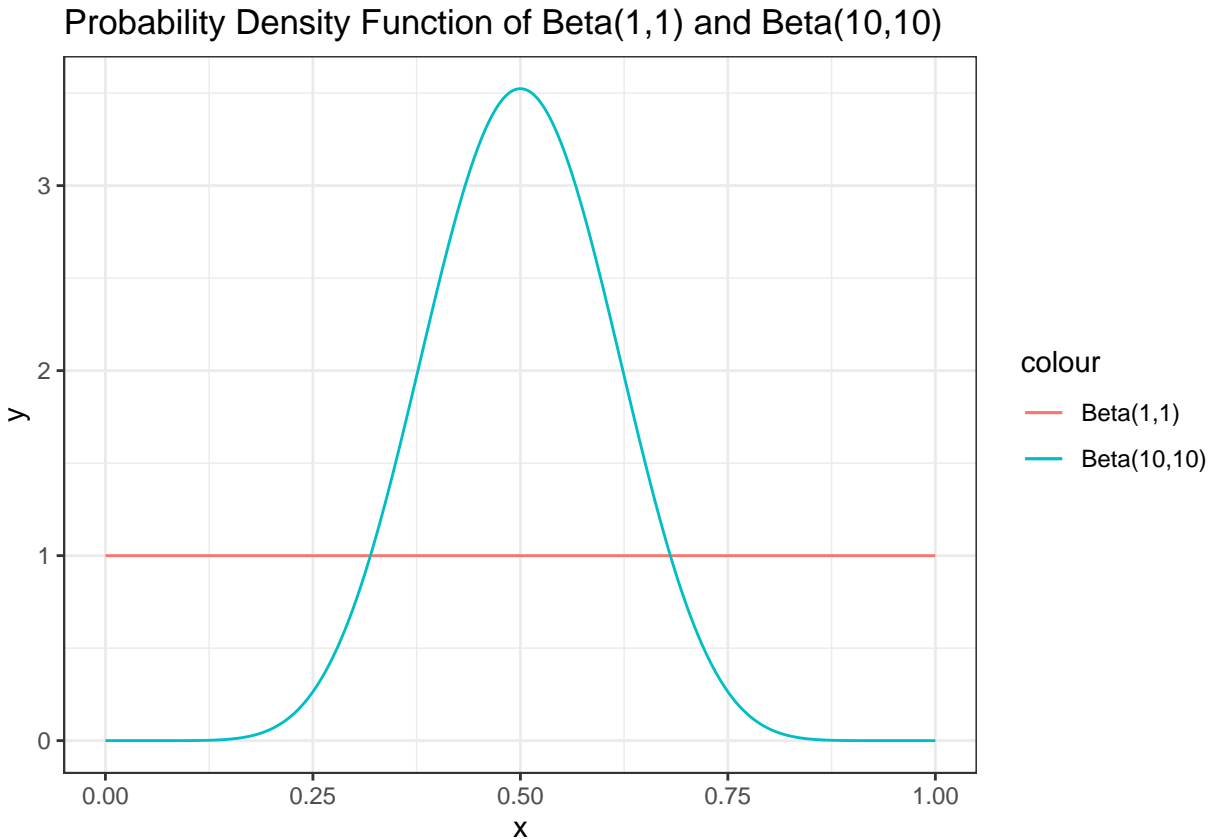
```
## Warning: package 'dplyr' was built under R version 4.2.2
```

```
## Warning: package 'forcats' was built under R version 4.2.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df = data.frame(x, y_1, y_2)

ggplot(df, aes(x)) +
  geom_line(aes(y = y_1, color = "Beta(1,1)")) +
  geom_line(aes(y = y_2, color = "Beta(10,10)")) +
  ggtitle("Probability Density Function of Beta(1,1) and Beta(10,10)") +
  xlab("x") +
  ylab("y") +
  theme_bw()
```



We can see that Beta(1,1) is uniformly distributed from 0 to 1 as we have no prior information so any  $\theta$  is equally likely. With Beta(10,10), we see that the probability density function is concentrated on 0.5, since from prior information of 9 success and 9 failure, we're more certain that  $\theta$  is around 0.5.

## 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 data
pt = 118

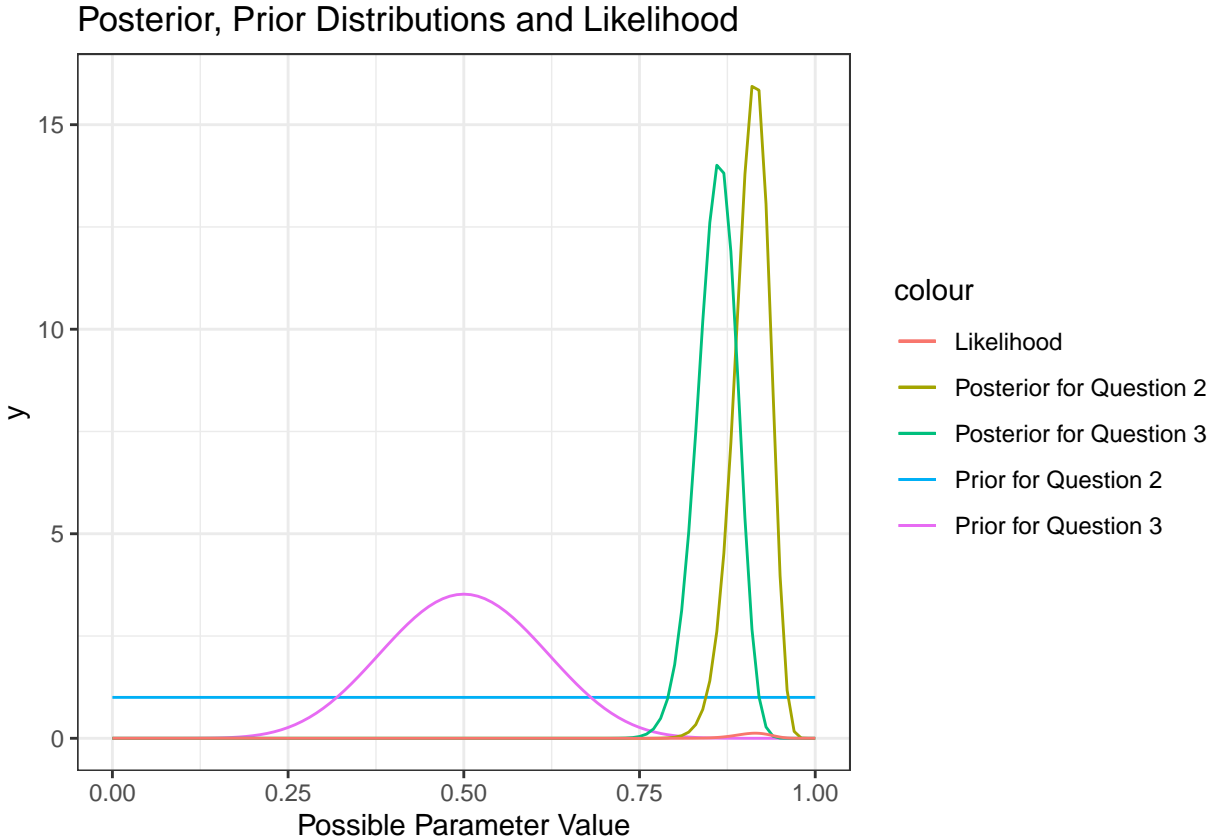
#possible theta values
p = seq(from = 0, to =1, by = 0.001)

lhood = dbinom(pt, size = 129, p)

x = data.frame(x = p, y =lhood)

x = seq(from = 0, to =1, by = 0.001)
x_data = data.frame(x, lhood = lhood)
ggplot(x_data,aes(x))+
  stat_function(fun = dbeta, args = list(shape1 = 1, shape2 = 1),
               aes(color = "Prior for Question 2"))+
  stat_function(fun = dbeta, args = list(shape1 = 10, shape2 = 10),
               aes(color = "Prior for Question 3"))+
  stat_function(fun = dbeta, args = list(shape1 = 119, shape2 = 12),
               aes(color = "Posterior for Question 2"))+
  stat_function(fun = dbeta, args = list(shape1 = 128, shape2 = 21),
               aes(color = "Posterior for Question 3"))+
  geom_line(aes(y = lhood, color = "Likelihood"))+
  labs(title= "Posterior, Prior Distributions and Likelihood")+
  ylab("y")+
  xlab("Possible Parameter Value")+
  theme_bw()

```



As expected, the likelihood peaks around 0.9 (and we know by MLE estimation that it peaks at 118/129), since our data suggests such high  $\theta$ .

The posterior distribution for question 2 is  $\text{Beta}(119, 12)$ , and the posterior distribution for question 3 is proportional to  $\theta^{118}(1 - \theta)^{11}\theta^9(1 - \theta)^9 = \theta^{127}(1 - \theta)^{20}$ , which means the posterior distribution will be  $\text{Beta}(128, 21)$ .

We see that with question 3 having a more informative prior that has a peak at 0.5, the posterior distribution peaks closer to 0.5 compared to the posterior distribution of question 2. Since in the prior for question 3, we had some certainty that  $\theta$  was around 0.5, the posterior distribution will have a peak closer to 0.5 compared to the posterior distribution of question 2.

## 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

First, note that  $\theta$  can range from  $-1$  to  $1$ .

A noninformative prior would be an uniform distribution over  $(-1, 1)$ , since it offers no information except for the possible range of  $\theta$ .

A subjective prior would be a distribution on  $(-1, 1)$  with the probability density function  $1 - |x|$ , which assigns smaller probability on values closer to  $-1$  or  $1$  as drastic improvement or drastic regression in the shooting ability is unlikely, and bigger probability on values close to  $0$  as we are being conservative in estimating  $\theta$  and assuming that there is likely no improvement.