

# Week 3 Lab: Intro to Bayes

J. Arturo Esquivel

25/01/23

## 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.

$Y$ , the number of women indicating happiness, follows a  $Bin(n, \theta)$ . Thus, its log-likelihood function is given by

$$l(y; \theta) \propto y \log(\theta) + (n - y) \log(1 - \theta).$$

Differentiating with respect to  $\theta$ , and equating to zero yields that  $\hat{\theta}_{MLE} = y/n = 118/129 = 0.91$ . Now, for a large value of  $n$ ,  $\hat{\theta}_{MLE} \sim N(\mu, \sigma^2)$ . Where  $\mu = E[y/n] = \theta/n$  and  $\sigma^2 = Var(y/n) = \theta(1 - \theta)/n$ . Approximating both quantities by substituting  $\hat{\theta}_{MLE}$  we can get a Wald 95% CI as  $\hat{\mu} \pm 1.96\hat{\sigma}^2 = (0.87, 0.96)$ .

## Question 2

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

A  $Beta(1,1)$  distribution is equal to 1. Hence

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

Which corresponds to the kernel of a  $Beta(y + 1, n - y + 1)$ . Thus the posterior mean of  $\theta$  is

$$\frac{y + 1}{n + 2} = \frac{119}{131} = 0.91.$$

For the credible interval, we can use the the interval given by the 2.5% and 97.5% quantiles of a  $Beta(119, 12)$ . Such interval is (0.85, 0.95).

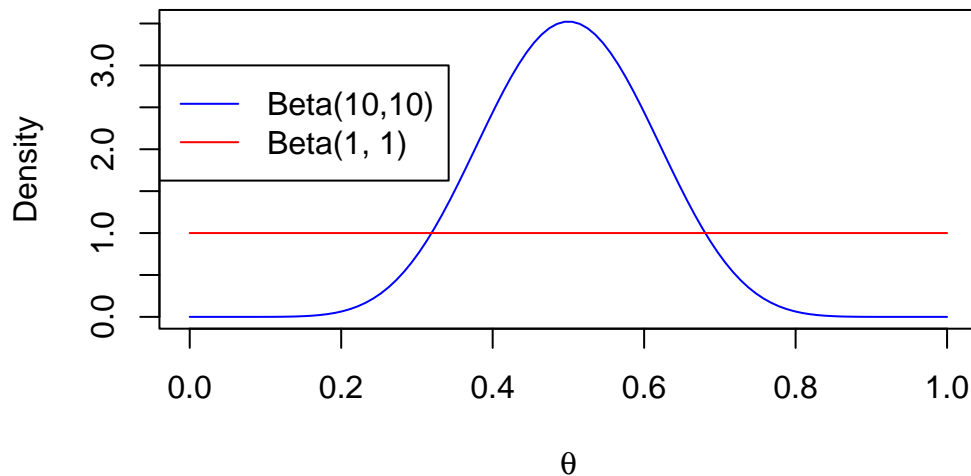
### Question 3

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?

The plot below shows both priors. As shown above (as well as in the plot), the  $\text{Beta}(1,1)$  is a noninformative prior, whereas the  $\text{Beta}(10,10)$  provides some information. The latter can be associated with a binomial distribution from which 9 successes and 9 failures have occurred. Thus providing information on the most likely vicinity of the distribution's parameter. Since the number of successes and failures is the same, the distribution is centered around 0.5. Holding much more information than the  $\text{Beta}(1,1)$ .

```
#define range
p = seq(0,1, length=100)

plot(p, dbeta(p, 10, 10), ylab='Density', xlab = bquote(theta), type = 'l', col='blue')
lines(p, dbeta(p, 1, 1), col='red')
legend(-.05, 3,
      c('Beta(10,10)', 'Beta(1, 1)'),
      lty=c(1,1), col = c('blue', 'red')
    )
```



## Question 4

Create a graph in ggplot which illustrates

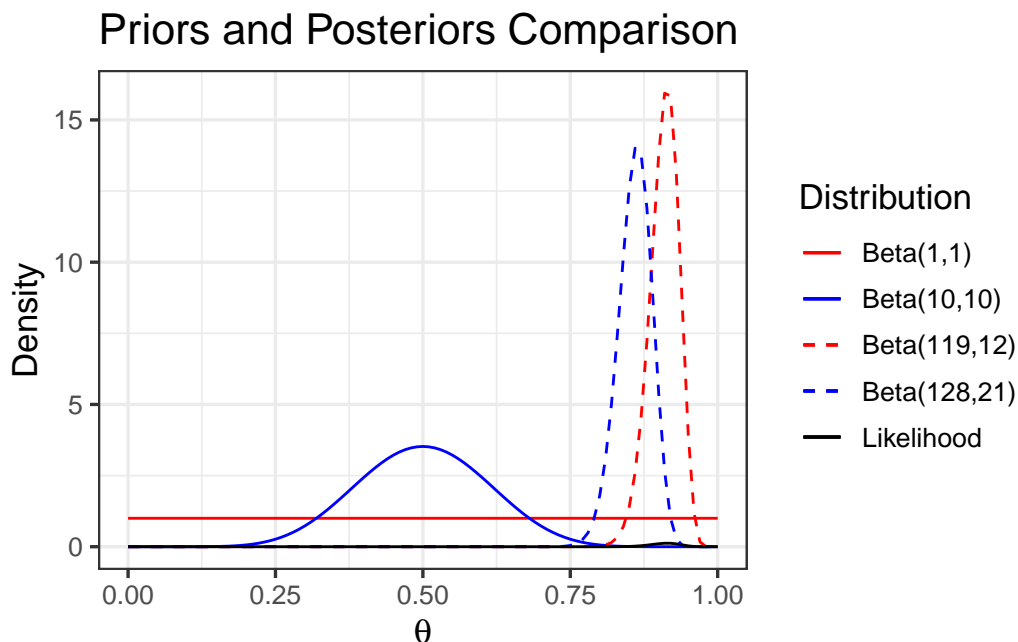
- The likelihood (easiest option is probably to use `geom_histogram` to plot the histogram of appropriate random variables)
- The priors and posteriors in question 2 and 3 (use `stat_function` to plot these distributions)

Comment on what you observe.

The plot is shown below. Solid (colored) lines correspond to priors and dashed lines to the posterior associated with the prior of the same color. As shown in Q3, the prior from Q3 provides some information as compared to the one from Q2. The likelihood (black line) reaches its maximum around 0.9, so both posteriors are placed close to that value, and without large dispersion. Note however, that the posterior from Q3 is to the left of the likelihood, because its prior is informative and centered around 0.5.

```
library(ggplot2)

lik <- function(x){choose(129,118)*(x^118)*(1-x)^11}
ggplot(data.frame(x = c(0, 1))) +
  xlim(c(0, 1)) +
  stat_function(
    fun = dbeta, args=list(1, 1), aes(colour="Beta(1,1)", linetype="Beta(1,1)")) +
  stat_function(
    fun = dbeta, args=list(10, 10), aes(colour="Beta(10,10)", linetype="Beta(10,10)")) +
  stat_function(
    fun = dbeta, args=list(119, 12), aes(colour="Beta(119,12)", linetype="Beta(119,12)")) +
  stat_function(
    fun = dbeta, args=list(128, 21), aes(colour="Beta(128,21)", linetype="Beta(128,21)")) +
  geom_function(
    fun = lik, aes(colour = "Likelihood", linetype = "Likelihood")) +
  scale_color_manual(
    "Distribution", values = c("red", "blue", "red", "blue", "black")) +
  scale_linetype_manual(
    "Distribution", values = c("solid", "solid", "dashed", "dashed", "solid")) +
  labs(
    x = bquote(theta), y = "Density", title = "Priors and Posteriors Comparison") +
  theme_bw(base_size = 13)
```



### 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
- Note that  $\theta$  goes from -1 (if all students go from making all shots to making none) to 1 (if all students go from making no shot to making all). Thus a logical noninformative prior for  $\theta$  is a  $Unif(-1, 1)$ , which covers the whole range and assigns the same probability to each possible value in it.
- For an informative prior we could use a  $N(0.25, 0.073)$ . Since the students practice for a month, it should be expected for their performance to improve (an average improvement

of 25% of the shots seems reasonable), thus  $\theta$  should assign most of the probability to positive values of  $\theta$  and practically zero to values outside of  $[-1,1]$ .