Lab 1: Mixture of Beta Priors

Author: Shashank Sule

Total Grade for Lab 1: /15

Comments (optional)

Template for lab report

Instructions: This is the template you will use to type up your responses to the exercises. To produce a document that you can print out and turn in just click on Knit PDF above. All you need to do to complete the lab is to type up your BRIEF answers and the R code (when necessary) in the spaces provided below.

It is strongly recommended that you knit your document regularly (minimally after answering each exercise) for two reasons.

1. Ensure that there are no errors in your code that would prevent the document from knitting.
2. View the instructions and your answers in a more legible, attractive format.

```
# Any text BOTH preceded by a hashtag AND within the ```{r} ``` code chunk is a commen
# R indicates a comment by turning the text green in the editor, and brown in the knit
# document.
# Comments are not treated as a command to be interpreted by the computer.
# They normally (briefly!) describe the purpose of your command or chunk in plain Engl
# However, for this class, they will have a different goal, as the text above and belo
# each chunk should sufficiently describe the chunk's contents.
# For this class, comments will be used to indicate where your code should go, or to g
# hints for what the code should look like.
```

Mixture of Beta Priors

Estimate the probability p of teen recidivism based on a study in which there were $n = 43$ individuals released from incarceration and $y = 15$ re-offenders within 36 months.

Exercise 1: Using a Beta$(2, 8)$ prior for p, plot the prior $\pi(p)$ and the posterior $\pi(p \mid y)$ as functions of p. Find the posterior mean and standard deviation of p. Find a 95% quantile-based credible interval. You can use either the exact solution or approximation through Monte Carlo simulation.

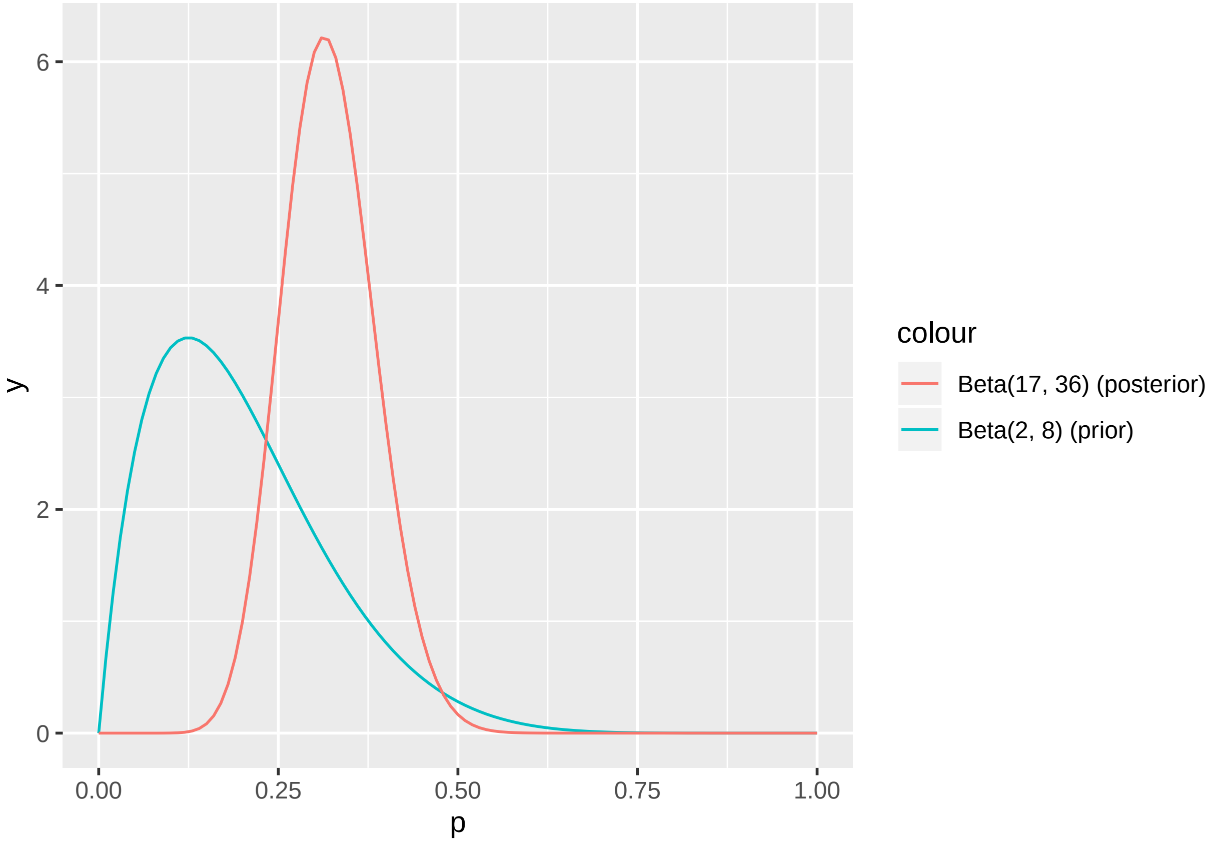
Using a Beta prior and a Binomial likelihood, the posterior distribution (from the result in class) is $\text{Beta}(2 + 15, 8 + 43 - 15) = \text{Beta}(17, 36)$

```
# If you use the exact solution...
# Here is some sample script to plot multiple Beta densities on the same graph
# Delete "eval=FALSE" above to see output

require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
ggplot(data = data.frame(p = c(0, 1)), aes(p)) +
  stat_function(fun = dbeta, args = list(shape1 = 2, shape2 = 8), aes(color = "Beta(2, 8
  stat_function(fun = dbeta, args = list(shape1 = 17, shape2 = 36), aes(color = "Beta(17
```




```
  ylab("Density")
```

```
## $y
## [1] "Density"
##
```

```
## attr(,"class")
## [1] "labels"
```

```r
# If you use approximation through Monte Carlo simulation...
# Here is some sample script to generate Beta samples and
# plot multiple Beta densities on the same graph
# Delete "eval=FALSE" above to see output
require(reshape2)
require(ggplot2)

set.seed(123)
S <- 100000
Beta11samples <- rbeta(S, shape1 = 1, shape2 = 1)
Beta55samples <- rbeta(S, shape1 = 5, shape2 = 5)
Beta46samples <- rbeta(S, shape1 = 4, shape2 = 6)
Beta64samples <- rbeta(S, shape1 = 6, shape2 = 4)
df <- as.data.frame(cbind(seq(1:S), Beta11samples, Beta55samples, Beta46samples, Beta64s
names(df) <- c("Index", "Beta(1, 1)", "Beta(5, 5)", "Beta(4, 6)", "Beta(6, 4)")
df_long <- melt(df, id = "Index")

ggplot(data = df_long, aes(value, colour = variable)) +
  geom_density() +
  xlab("p") + ylab("Density")
```

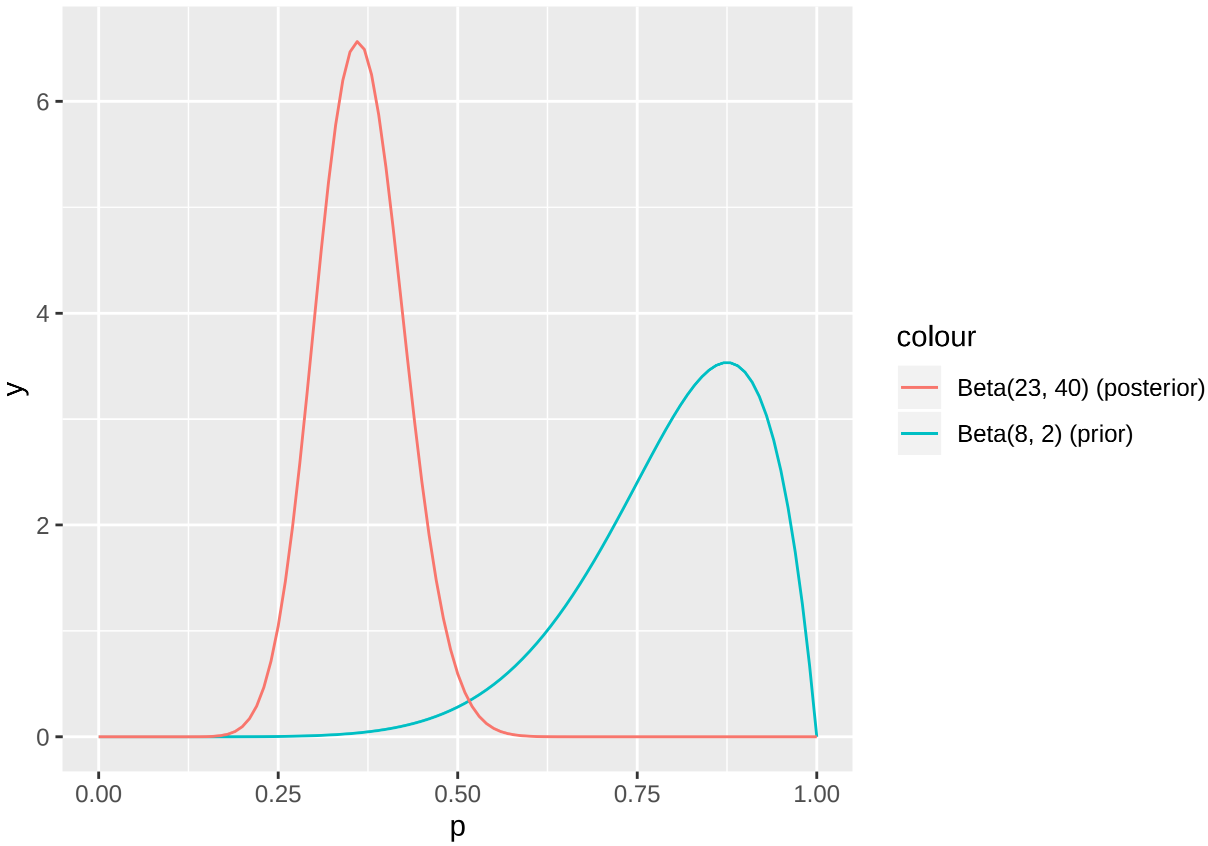
Grade for Exercise 1: /5

Comments:

Exercise 2: Repeat Exercise 1, but using a Beta$(8, 2)$ prior for p.

In this case, the posterior will be $\text{Beta}(8 + 15, 2 + 43 - 15) = \text{Beta}(23, 40)$

```r
ggplot(data = data.frame(p = c(0, 1)), aes(p)) +
  stat_function(fun = dbeta, args = list(shape1 = 8, shape2 = 2), aes(color = "Beta(8, 2
  stat_function(fun = dbeta, args = list(shape1 = 23, shape2 = 40), aes(color = "Beta(23
```




```r
  ylab("Density")
```

```
## $y
## [1] "Density"
##
## attr(,"class")
## [1] "labels"
```

Grade for Exercise 2: /5

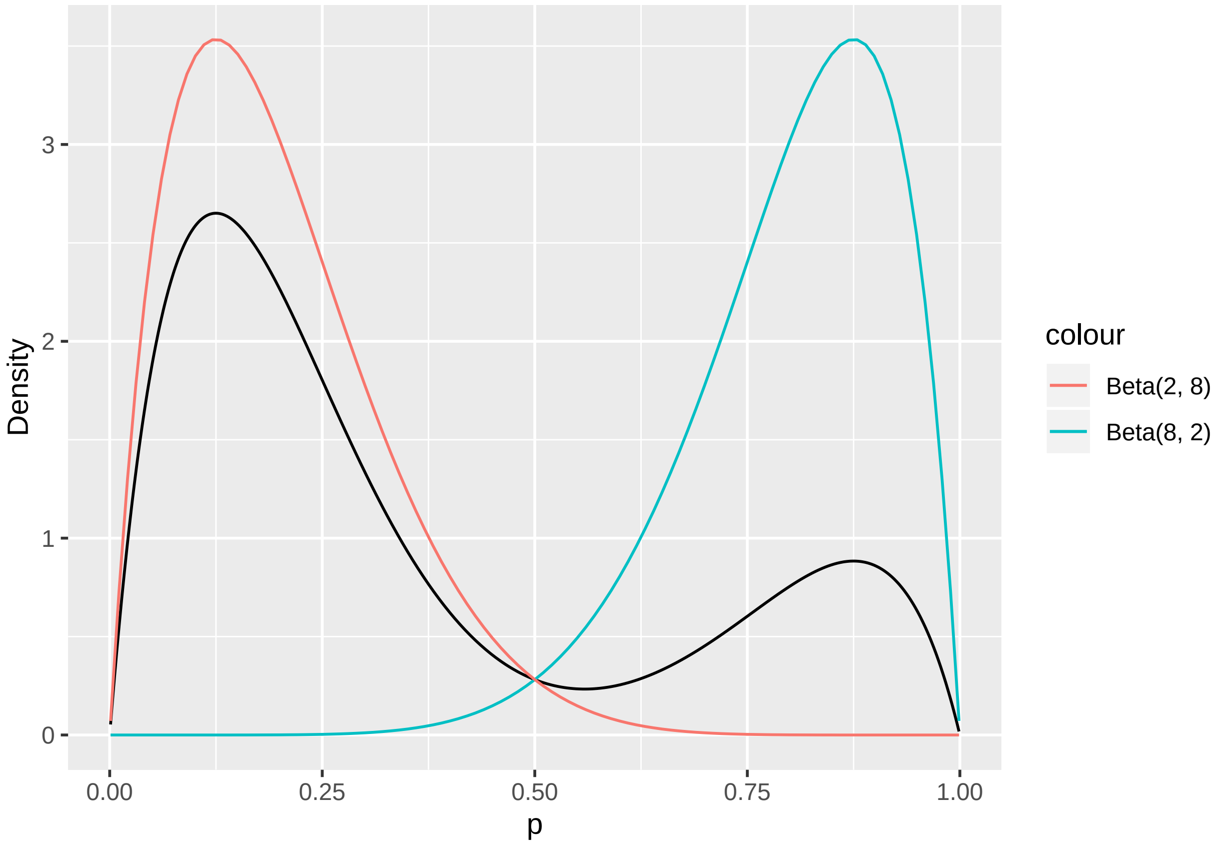
Comments:

Exercise 3: Consider the following prior distribution for p, a $75 - 25\%$ mixture of a Beta$(2, 8)$ and a Beta$(8, 2)$ prior distribution. Plot this prior distribution and compare it to the priors in Exercise 1 and Exercise 2. Describe what sort of prior opinion this may represent.

$$\pi(p) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}[3p(1-p)^7 + p^7(1-p)],$$

```
# If you use the exact solution...
# Here is some sample script to create density of a 75%-25% mixture Beta prior
# Delete "eval=FALSE" above to see output

p <- seq(0.001, 0.999, length = 1000)
mixture <- 0.75*dbeta(p, 2, 8) + 0.25*dbeta(p, 8, 2)
mixdis = data.frame(p,mixture)
ggplot(mixdis, aes(p,mixture)) +
  geom_line() +
  ylab("Density") +
  stat_function(fun = dbeta, args = list(shape1 = 8, shape2 = 2), aes(color = "Beta(8, 2
  stat_function(fun = dbeta, args = list(shape1 = 2, shape2 = 8), aes(color = "Beta(2, 8
```




```
  ylab("Density")
```

```
## $y
## [1] "Density"
##
## attr(,"class")
## [1] "labels"
```

```
# If you use approximation through Monte Carlo simulation...
# Here is some sample script to generate draws of a 75%-25% mixture Beta prior
```

```r
# Delete "eval=FALSE" above to see output

set.seed(123)
S <- 1000
MixtureBetaSamples <- rep(NA, S)
for (s in 1:S){
  component <- rbernoulli(1, 0.75)
  if (component == 1){
    MixtureBetaSamples[s] <- rbeta(1, shape1 = 1, shape2 = 1)
  }
  else
    MixtureBetaSamples[s] <- rbeta(1, shape1 = 5, shape2 = 5)
}
```

The mixture represents the belief that p is bimodally distributed with the two modes being approximately 0.125 and 0.875, with a bigger belief weighted towards 0.125. It's akin to saying "It's either 0.875 or 0.125 and I'm 75% sure it's 0.125"

Grade for Exercise 3: /5

Comments: