Homework #2

Dustin Leatherman

9/28/2020

2

Write an R function to perform a Bayesian Analysis of count data. The response, the number of successes in N independent trials, is $Y \in \{0, 1, 2, ..., N\}$ and the unknown parameter is the success of probability, $\theta \in [0, 1]$. Assume the Bayesian model with likelihood $Y|\theta \sim Bin(N, \theta)$, prior $\theta \sim Beta(a, b)$, and thus posterior $\theta|Y \sim Beta(Y + a, N - Y + b)$

Write an R function that takes Y, N, a, and b as inputs. The function should produce a plot (clearly labeled!) that overlays the prior and posterior density functions (both using the dbeta function), and it should return a list with posterior mean and posterior standard deviation.

```
betabin <- function(Y, N, a, b) {
  library(tidyverse)

  theta <- rbeta(N, a, b)
  post.a <- Y + a
  post.b <- N - Y + b
  post <- rbeta(N, post.a, post.b)

  post_avg <- mean(post)
  post_sd <- sd(post)

  post.dist <- dbeta(post, post.a, post.b)
  prior.dist <- dbeta(theta, a, b)

  # create homogenous dataframes so they can be unioned
  df.post <- data.frame(x = post, y = post.dist, distribution = "posterior")
  df.prior <- data.frame(x = theta, y = prior.dist, distribution = "prior")
  p1 <-
    df.post %>%
      union(df.prior) %>%
        ggplot(aes(x = x, color = distribution, y = y)) +
          geom_line() +
          labs(x = bquote(theta), y = "Density", color = "Distribution") +
          xlim(0,1)
  ls <- list("mean" = post_avg, "sd" = post_sd, "plot" = p1)
  return (ls)
}
```

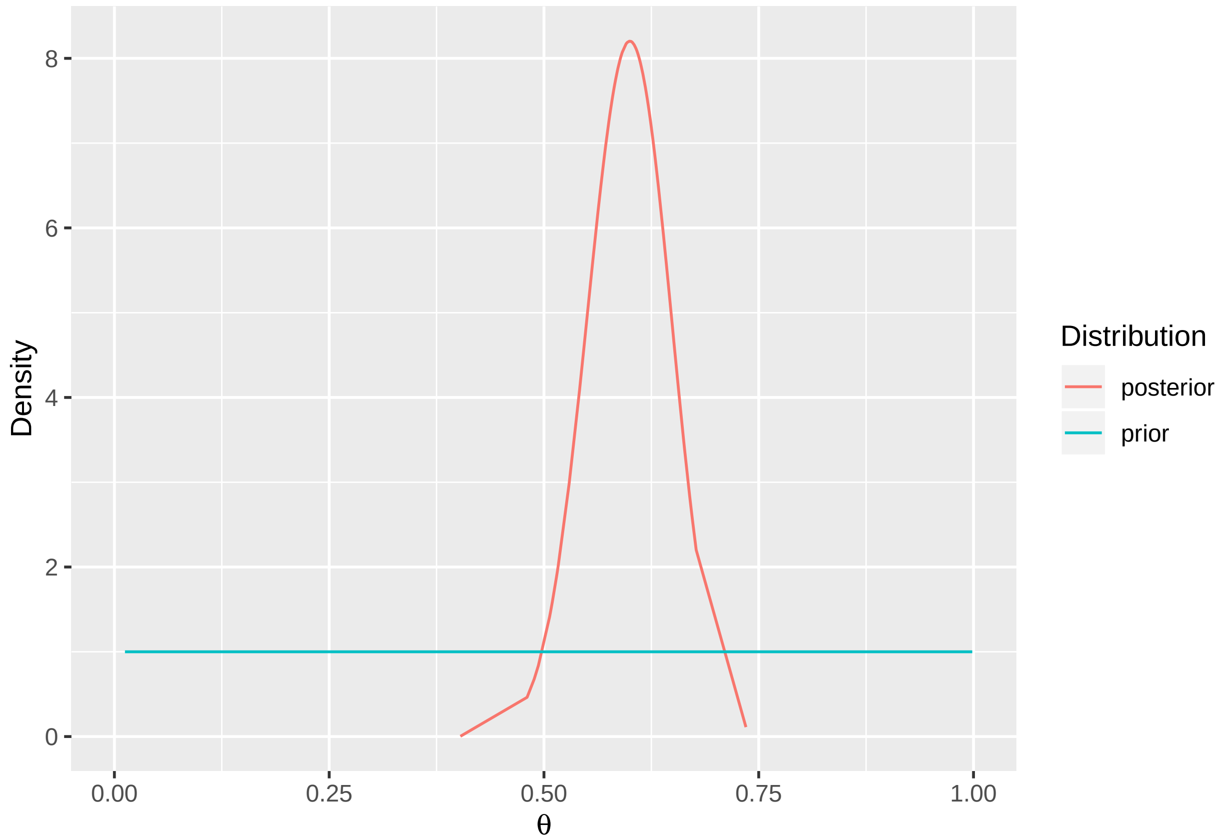
a

What values of a and b would make good default values to represent a prior that carries little information about θ? Make these the default values in your function

Beta(a = 1, b = 1) is equivalent to a Uniform(1,1) distribution which is considered an uninformative prior.

```r
betabin(60, 100, 1, 1)
```

```
## $mean
## [1] 0.5974314
##
## $sd
## [1] 0.05022078
##
## $plot
```



b

What values of a and b would give prior mean 0.7 and prior standard deviation 0.2? (**Hint**: Use the beta.prior.R program in Week Three of Course Content)

```r
source("~/Downloads/beta.prior.R")
res <- beta.prior(mean = 0.7, sd = 0.2)
```

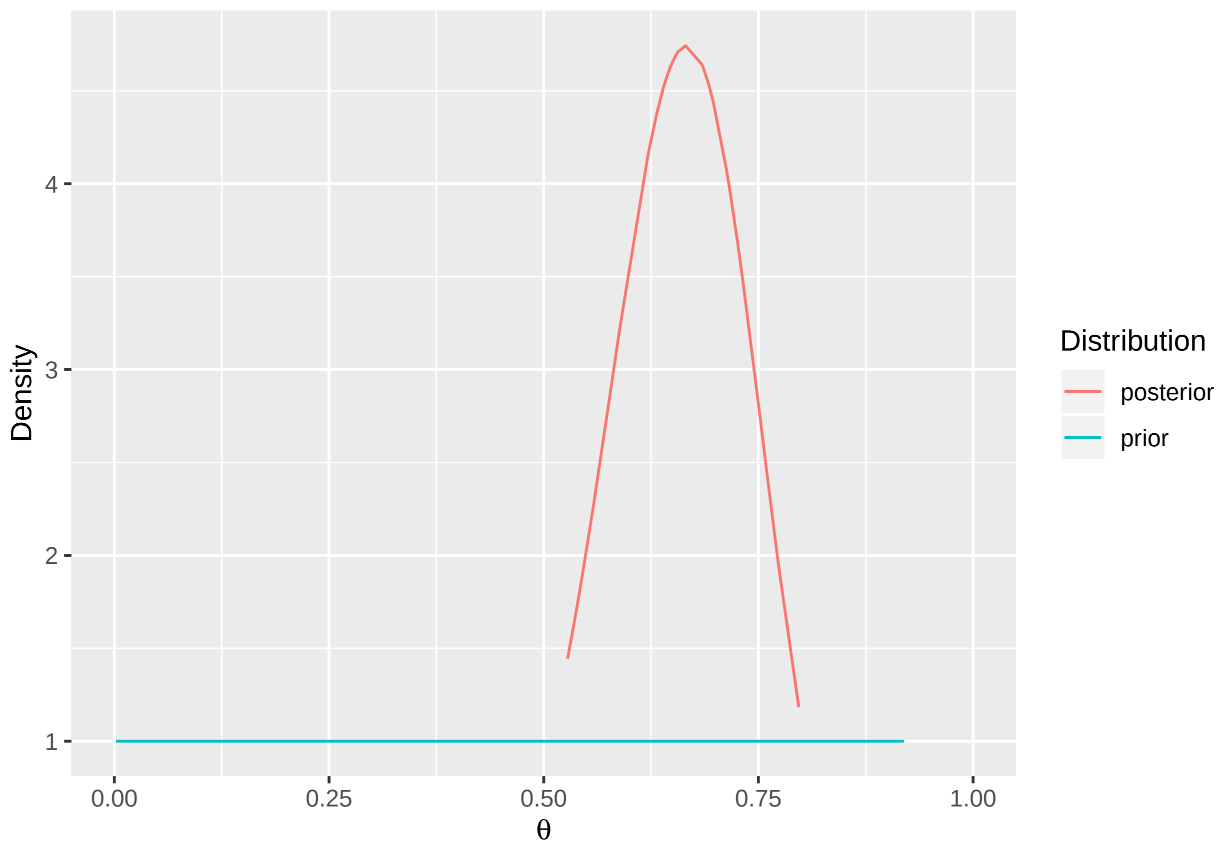
a = 2.975 and b = 1.275 would give a prior mean of 0.7 and prior standard deviation of 0.2.

c

Now we observe Y = 20 events in N = 30 trials. Use your code from #1 to conduct a Bayesian Analysis of these data. Perform the analysis twice: once with the uninformative prior from #2 and once with the informative prior from #3.

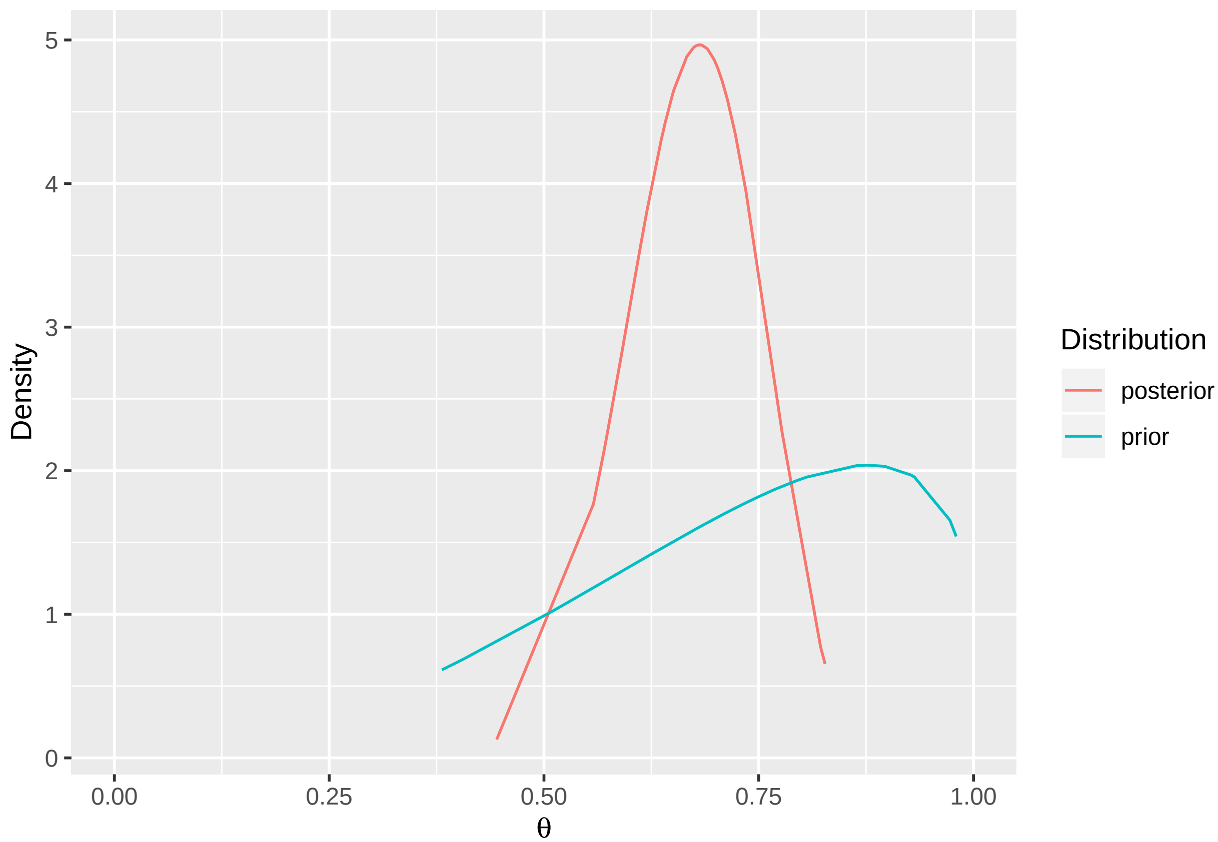
```r
betabin(20, 30, 1, 1)
```

```
## $mean
## [1] 0.6657758
##
## $sd
## [1] 0.07672714
##
## $plot
```




```r
betabin(20, 30, res$a, res$b)
```

```
## $mean
## [1] 0.6645733
##
## $sd
## [1] 0.07810849
##
## $plot
```



d

> Summarize the results. In particular, how does this anlaysis compare to a frequentist analysis and how much are the results affected by the prior?

The posterior in the first analysis with an uninformative prior has a slightly smaller mean, larger standard deviation, and a max value of approximately 4.75. Since the prior is uninformative, the posterior is close to the result of a MLE for a Beta Random Variable.

The second analysis shows the prior pulling the posterior slightly to the right and downwards. There are more values of θ covered by the posterior distribution using the informative prior; howeve, the effect of the informative prior appears to be negligible to the overall analysis.