STAT 447 Assignment 7

Caden Hewlett

2024-03-05

Question 1: Installing and Running Stan

Setting up the beta-binomial environment:

Here, we have observed data n and k.

We have a parameter p, where $p \sim \text{beta}(\alpha, \beta)$ and $k \sim \text{bin}(n, p)$.

```
data {
  int<lower=0> n;          // number of trials
  int<lower=0,upper=n> k; // number of successes
}

parameters {
  real<lower=0,upper=1> p; // p in [0, 1]
}

model {
  // prior
  p ~ beta(1,1);

  // likelihood
  k ~ binomial(n, p);
}
```

Then, we run the MCMC to find $\mathbb{P}(p \mid \{k, n\} = \{3, 3\})$. As in, the posterior success probability given three subsequent successes.

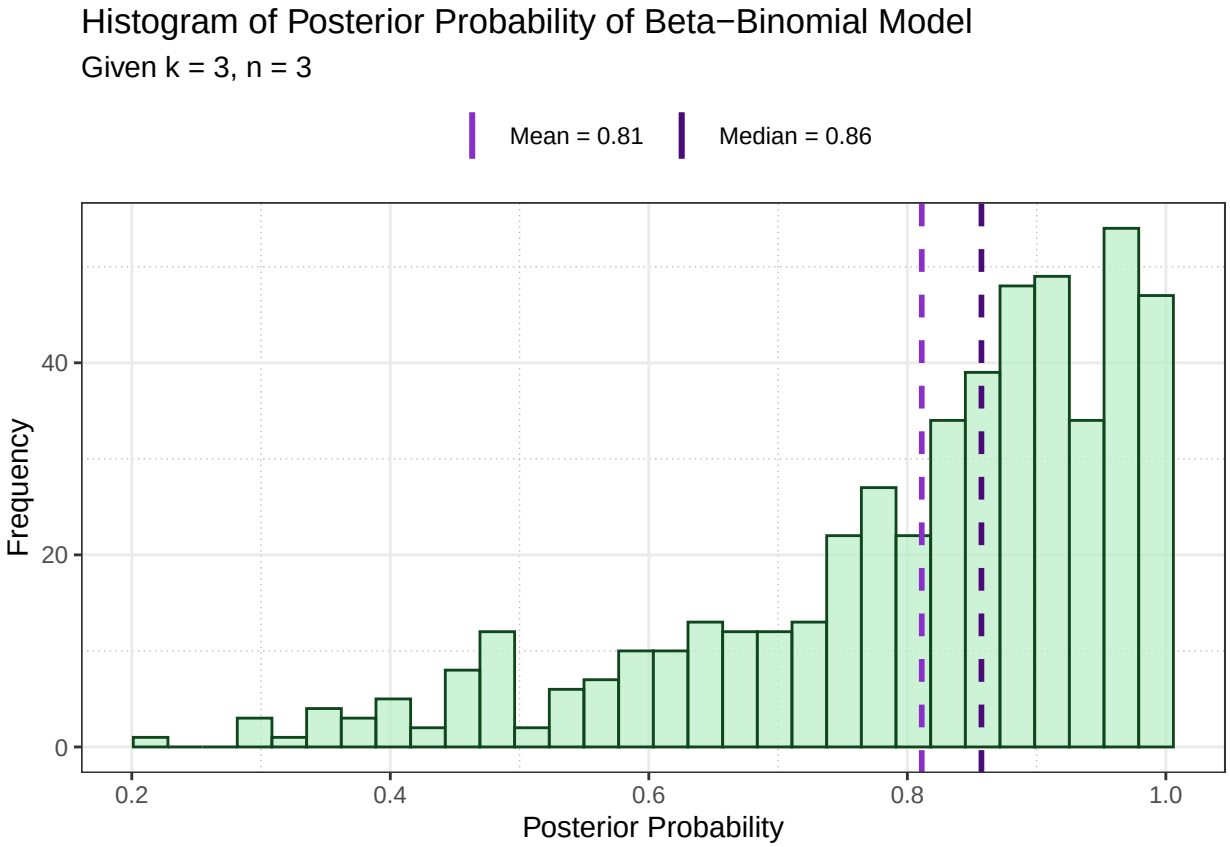
```
require(rstan)

fit = sampling(
  test,
  seed = 123,
  data = list(n = 3, k = 3),
  chains = 1,
  iter = 1000
)

q1model = extract(fit)
```

We can also use `ggplot2` to make a nice histogram of the output.

```r
ggplot(data.frame(q1model$p), aes(x = q1model.p)) +
  geom_histogram(
    bins = 30,
    fill = "#B7EFC5",
    color = "#10451D",
    alpha = 0.7
  ) +
  geom_vline(aes(xintercept = mean(q1model$p), color = "Mean"),
             linetype = "dashed",
             linewidth = 1) +
  geom_vline(
    aes(xintercept = median(q1model$p), color = "Median"),
    linetype = "dashed",
    linewidth = 1
  ) +
  scale_color_manual(
    name = "",
    values = c("Mean" = "#8B2FC9",
               "Median" = "#4A0A77"),
    labels = c(paste("Mean =", round(mean(q1model$p), 2)),
               paste("Median =", round(median(q1model$p), 2)))
  ) +
  labs(
    title = "Histogram of Posterior Probability of Beta-Binomial Model",
    subtitle = "Given k = 3, n = 3",
    x = "Posterior Probability",
    y = "Frequency"
  ) +
  theme_bw() +
  theme(
    legend.position = "top",
    panel.grid.minor = element_line(colour = "gray", linetype = "dotted")
  ) +
  guides(color = guide_legend(override.aes =
                                list(linetype = c("solid", "solid"))))
```



So, both from the histogram we can see the posterior median is approximately 0.86.

Precisely, it is the value below:

```
median(q1model$p)
```

```
## [1] 0.8573115
```