Stat 136 (Bayesian Statistics)

Lesson 1.2: Bayesian Inference for the Binomial Proportion: Discrete Prior

Three general steps of Bayesian inference:

1. Express an opinion about the location of the proportion p before sampling (**prior**).

2. Take the sample and record the observed proportion (**data/likelihood**).

3. Use Bayes' rule to sharpen or update the previous opinion about p given the information in the sample (**posterior**).

Example 1.2.1

Suppose a person is interested in learning about the sleeping habits of graduating college students. He hears that doctors recommend eight hours of sleep for an average adult. What proportion of graduating students get at least eight hours of sleep?

Let p be the proportion of all graduating students who sleep (on a typical night) at least eight hours. We are interested in learning about the location of p.

The value of the proportion p is unknown. In the Bayesian viewpoint, a person's beliefs about the uncertainty in this proportion is represented by a probability distribution placed on this parameter. This distribution reflects the person's subjective prior opinion about plausible values of p.

A random sample of graduating students from a particular university will be taken to learn about this proportion. But first the researcher does some initial research to learn about the sleeping habits of graduating college students. This research will help him in constructing a prior distribution.

The prior distribution for p

One article about sleep patterns of students presented that that college students generally get less than eight hours of sleep and so p (the proportion that sleep at least eight hours) is likely smaller than 0.5. After some reflection, her best guess at the value of p is .3. But it is very plausible that this proportion could be any value in the interval from 0 to 0.5.

A simple approach for assessing a prior for p is to write down a list of plausible proportion values and then assign weights to these values. The person in our example believes that

0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95

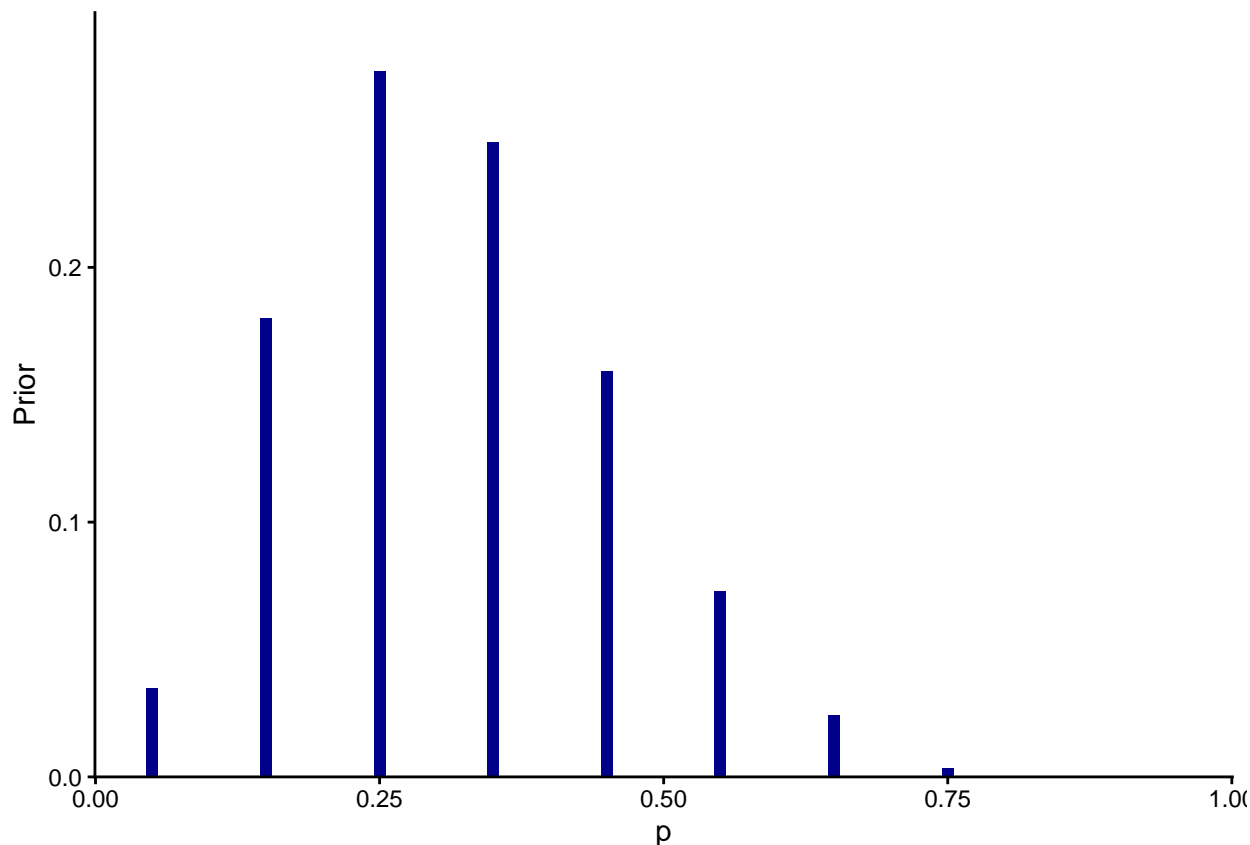are possible values for p. Based on his beliefs, he assigns these values the corresponding weights

1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0

which can be converted to prior probabilities by dividing each weight by their sum.

```r
library(tidyverse)
library(patchwork)
p <- seq(0.05, 0.95, by = 0.1)
wt <- c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
prior <- wt/sum(wt)
PriorD <- as.data.frame(cbind(p, prior))
knitr::kable(PriorD)
```

p	prior
0.05	0.0346021
0.15	0.1799308
0.25	0.2768166
0.35	0.2491349
0.45	0.1591696
0.55	0.0726644
0.65	0.0242215
0.75	0.0034602
0.85	0.0000000
0.95	0.0000000

```r
p1 <- PriorD %>%
  ggplot(aes(x = p, y = prior)) +
  geom_bar(stat = "identity", width = 0.01, fill="darkblue") +
  scale_y_continuous(expand=c(0,0), limit = c(0, 0.3)) +
  labs(y = "Prior")+
  theme_classic()
p1
```

The likelihood of the data

A sample of 27 graduating students is taken – in this group, 11 record that they had at least eight hours of sleep the previous night. Based on the prior information and these observed data, the researcher is interested in estimating the proportion p.

If we regard a "success" as sleeping at least eight hours and we take a random sample of size n with s successes and f failures, then the likelihood function is given by

$$L(y|p) = \binom{n}{s} p^s(1-p)^f \propto p^s(1-p)^f$$

In our example, 11 of 27 students sleep a sufficient number of hours, so $s = 11$ and $f = 16$, and the likelihood function is

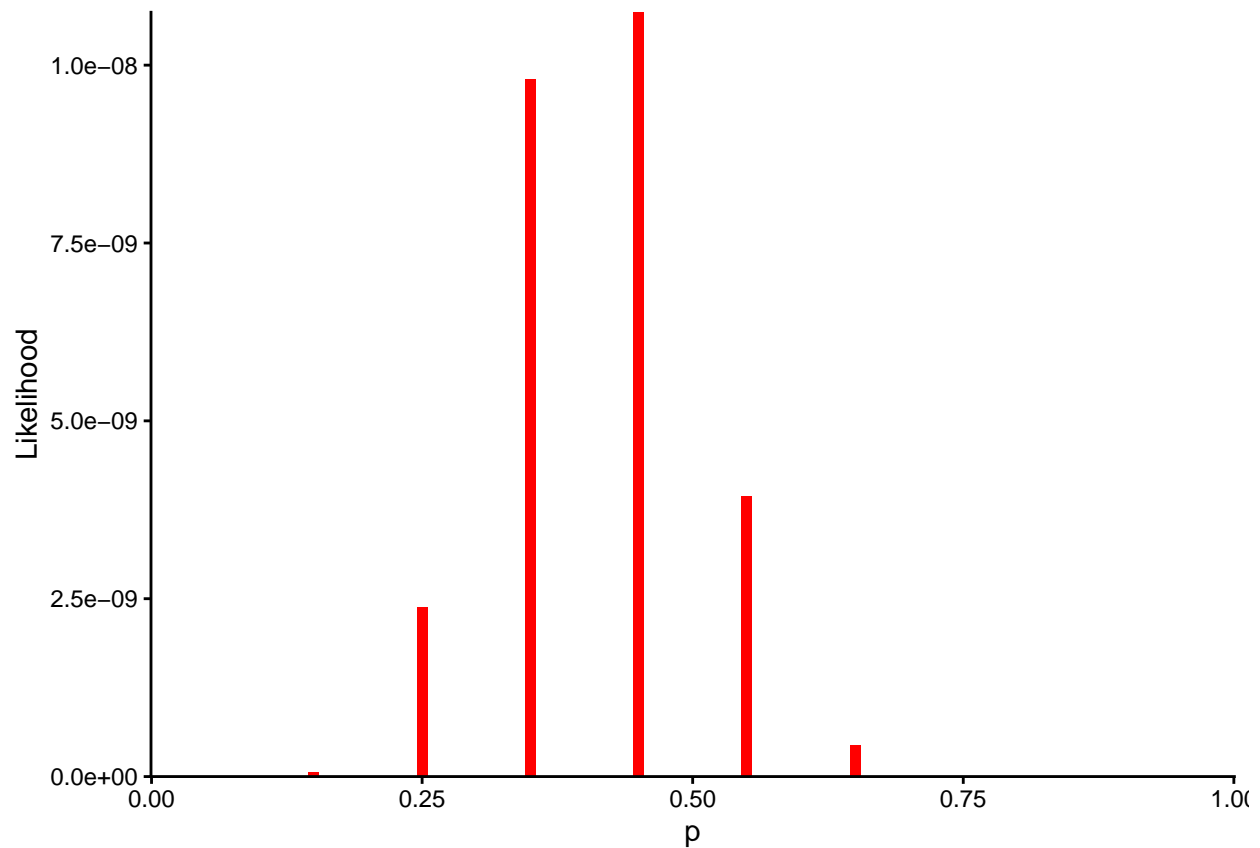$$L(y|p) \propto p^{11}(1-p)^{16}$$

```
likelihood <- p^{11}*(1-p)^{16}
p2 <- as.data.frame(cbind(p, prior,likelihood)) |>
  ggplot(aes(x = p,
             y = likelihood)) +
```

3

```
  geom_bar(stat = "identity", width = 0.01, fill="red") +
  scale_y_continuous(expand=c(0,0), limit = c(0, max(likelihood))) +
  labs(y = "Likelihood")+
  theme_classic()
#p2 <- plot(p, likelihood, type="h", ylab="Likelihood(p)")
p2
```


The posterior distribution

```
PL <- prior*likelihood
post <- round(PL/sum(PL),7)
cbind(p, prior, likelihood, PL, post)
```
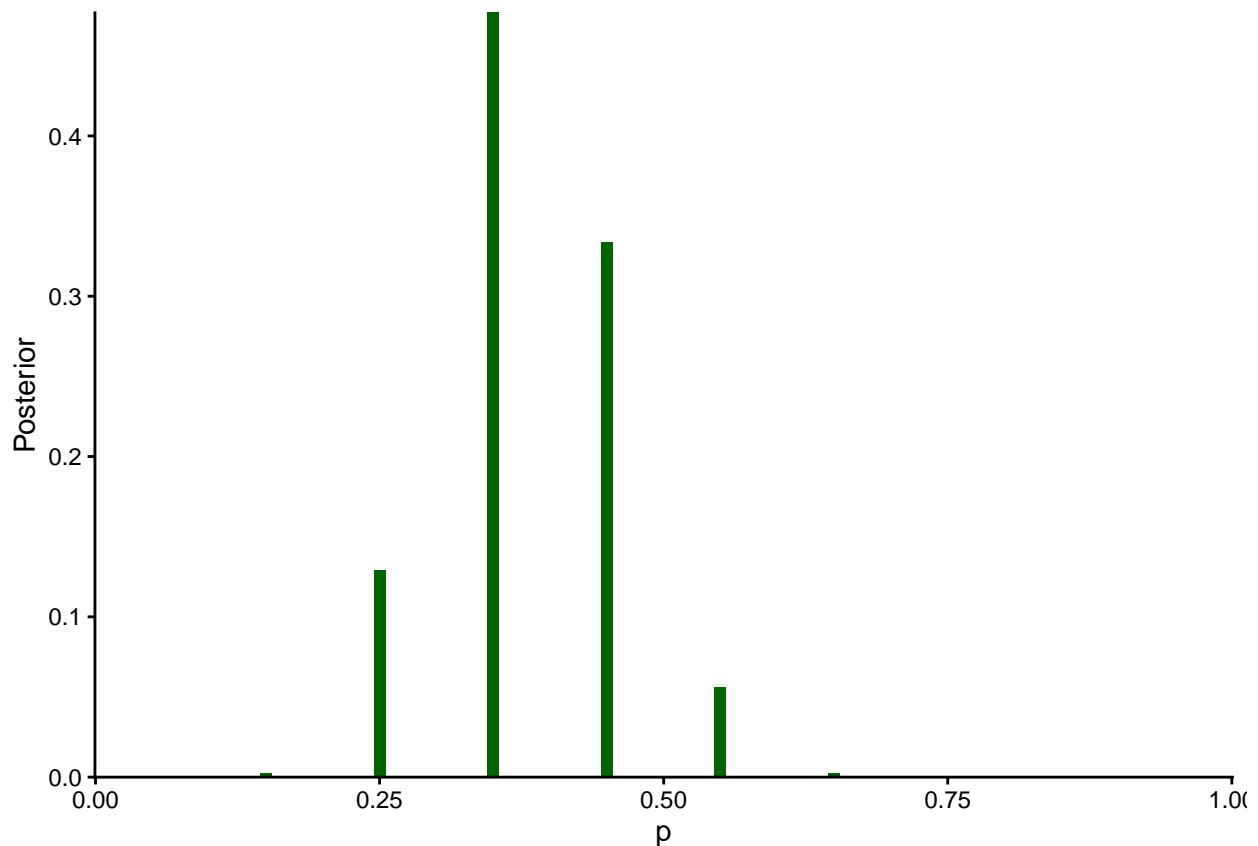
```
##          p        prior    likelihood          PL        post
## [1,] 0.05 0.034602076 2.149056e-15 7.436180e-17 0.0000000
## [2,] 0.15 0.179930796 6.422538e-11 1.155612e-11 0.0022560
## [3,] 0.25 0.276816609 2.389573e-09 6.614735e-10 0.1291349
## [4,] 0.35 0.249134948 9.803072e-09 2.442288e-09 0.4767910
```

4

```
##   [5,] 0.45 0.159169550 1.074337e-08 1.710018e-09 0.3338350
##   [6,] 0.55 0.072664360 3.939035e-09 2.862274e-10 0.0558782
##   [7,] 0.65 0.024221453 4.437472e-10 1.074820e-11 0.0020983
##   [8,] 0.75 0.003460208 9.833634e-12 3.402641e-14 0.0000066
##   [9,] 0.85 0.000000000 1.099179e-14 0.000000e+00 0.0000000
## [10,] 0.95 0.000000000 8.679201e-22 0.000000e+00 0.0000000
```

```r
print(sum(PL)) # To calculate the normalizing constant
```

```
## [1] 5.122345e-09
```
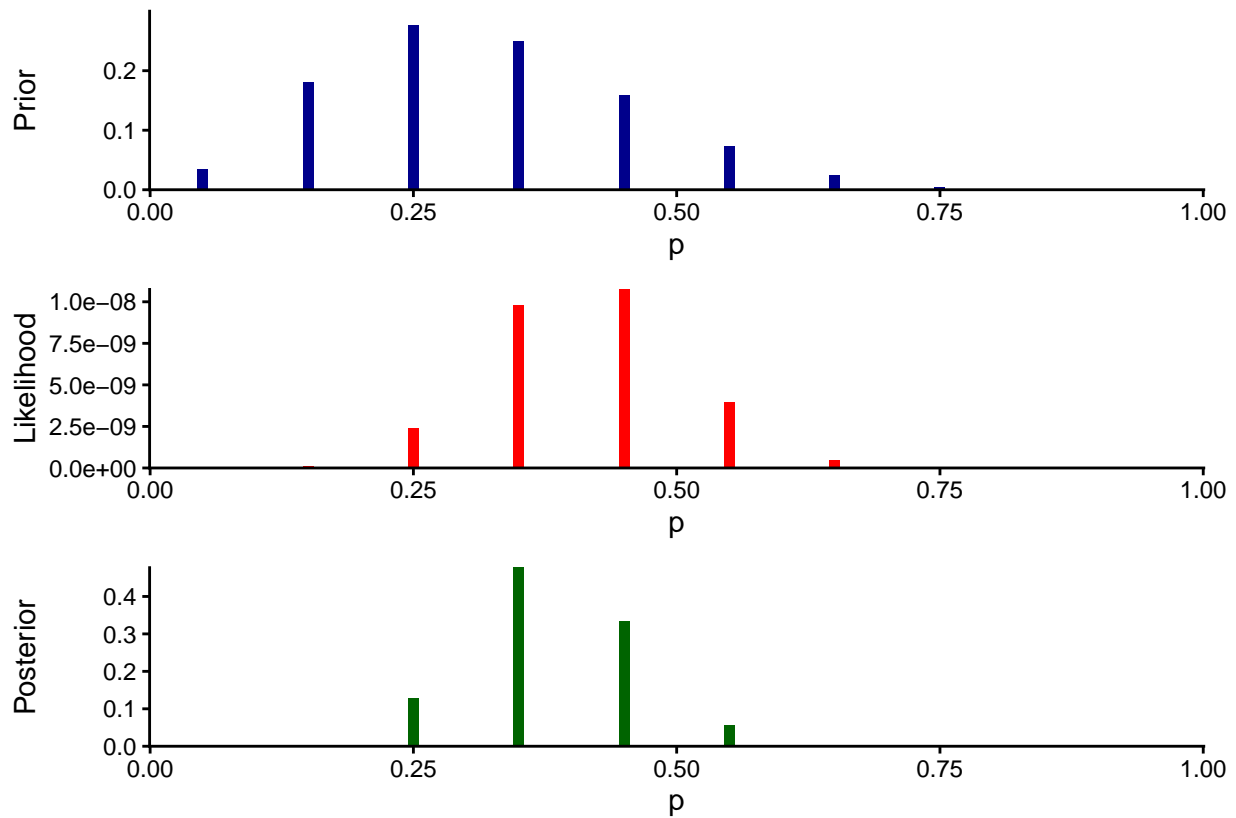
```r
p3 <- as.data.frame(cbind(p, prior, likelihood, PL, post)) |>
  ggplot(aes(x = p,
             y = post)) +
  geom_bar(stat = "identity", width = 0.01, fill="darkgreen") +
  scale_y_continuous(expand=c(0,0), limit = c(0, max(post))) +
  labs(y = "Posterior")+
  theme_classic()
p3
```


For better comparison, we can place the plots on top of each other.

```
p1 / p2 / p3
```


Better alternative

The R function *pdisc()* in the package **LearnBayes** computes the posterior probabilities. To use *pdisc()*, one inputs the vector of proportion values p, the vector of prior probabilities prior, and a data vector data consisting of s and f. The output of *pdisc()* is a vector of posterior probabilities.

```r
library(LearnBayes)
data <- c(s = 11, f = 16)
post1 <- pdisc(p, prior, data)
knitr::kable(as.data.frame(cbind(p,prior,post1)))
```

p	prior	post1
0.05	0.0346021	0.0000000
0.15	0.1799308	0.0022560
0.25	0.2768166	0.1291349

p	prior	post1
0.35	0.2491349	0.4767910
0.45	0.1591696	0.3338350
0.55	0.0726644	0.0558782
0.65	0.0242215	0.0020983
0.75	0.0034602	0.0000066
0.85	0.0000000	0.0000000
0.95	0.0000000	0.0000000

Inference questions

1. What is the expected proportion of graduating students who have slept at least eight hours per night?

2. What is the chance that between 30-50 percent of graduating students have slept at least eight hours per night?

3. Is there a statistical evidence that less than 50% of graduating students have slept at least eight hours per night?

To answer Question 1, we can use the classical estimator which is given by

$$\hat{p} = \frac{11}{27} \approx 0.41$$

About 41% of graduating students have slept at least eight hours per night.

In Bayesian statistics, there are many statistics which we can use to summarize the posterior distribution and come up with an estimate of p. One exampe is the mean of the posterior distribution. For this example, the posterior mean is

$$\hat{p}_{Bayes} \approx 0.38$$

While the mode of the posterior distribution is 0.35.

The answer to Question 2 is:

$$P(0.30 \leq p \leq 0.5) = P(p = 0.35) + P(p = 0.45) = 0.8106$$

The answer to Question 3 is some kind of a hypothesis test:

- $H_0 : p = 0.5$ vs $H_1 : p < 0.5$

$$\begin{aligned} P(H_1) &= P(p < 0.5) \\ &= P(p = 0.45) + P(p = 0.35) + P(p = 0.25) + P(p = 0.15) + P(p = 0.05) \\ &\approx 0.9420 \end{aligned}$$

Therefore, it is highly likely that less than 50% of graduating students have slept at least eight hours per night.

Question to ponder

What happens if the prior distribution assigns equal probability (**non-informative prior**) for each possible value of p?

Takeaways

- Bayesian inference (point estimation, credible interval, and hypothesis testing) is focused on the posterior distribution of the parameter p

- The posterior distribution of the parameter is determined by the prior distribution and the likelihood (data)

- If the prior distribution is "non-informative" the posterior distribution is "dominated" by the likelihood