STAT 339: Statistical Theory

Bayesian Hypothesis Testing

Anthony Scotina

Bayes Factors

A Slide about R

Suppose I am interested in estimating the **proportion**, p, of *students at Simmons who have heard of R*.

- A *colleague* claims that this proportion is **over 25%**, and I want to test how plausible this claim is.

I actually think p is over 1/3 but I'm not very confident in that estimate. Suppose we use the following hierarchy:

- **Prior**: $p \sim Beta(9, 18)$

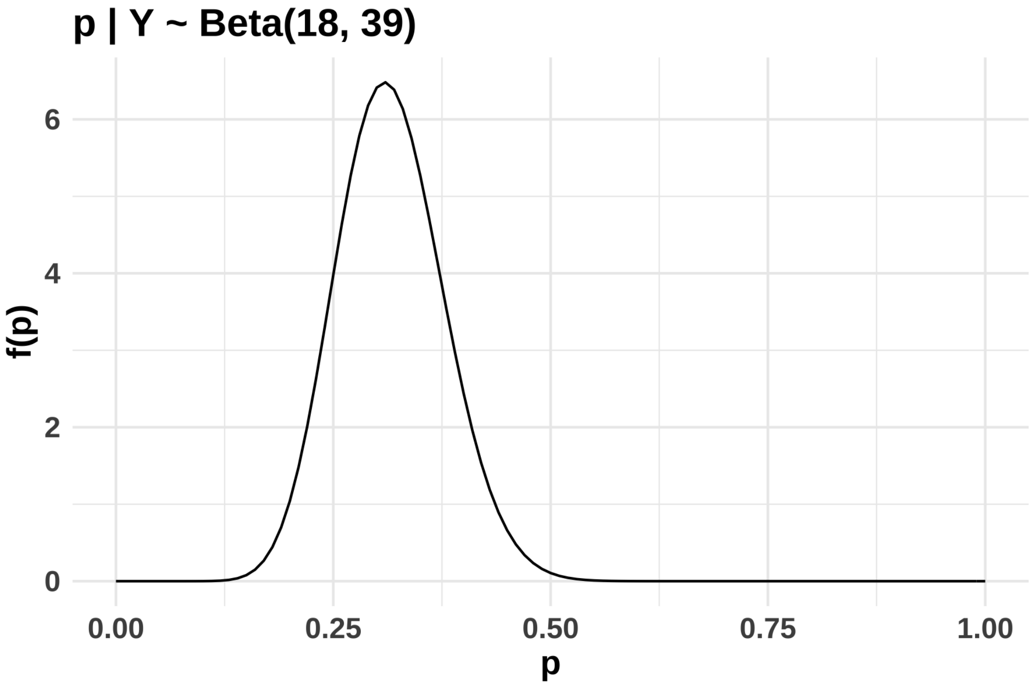
- **Data**: $Y \mid p \sim Binomial(n = 30, p)$, where we observe $Y = 9$

Using the *Beta-Binomial conjugacy*, our **posterior** distribution is

$$p \mid Y \sim Beta(9 + 9 = \mathbf{18}, 18 + 30 - 9 = \mathbf{39})$$

Is the "over 25%" claim plausible?

It's looking good! 😎

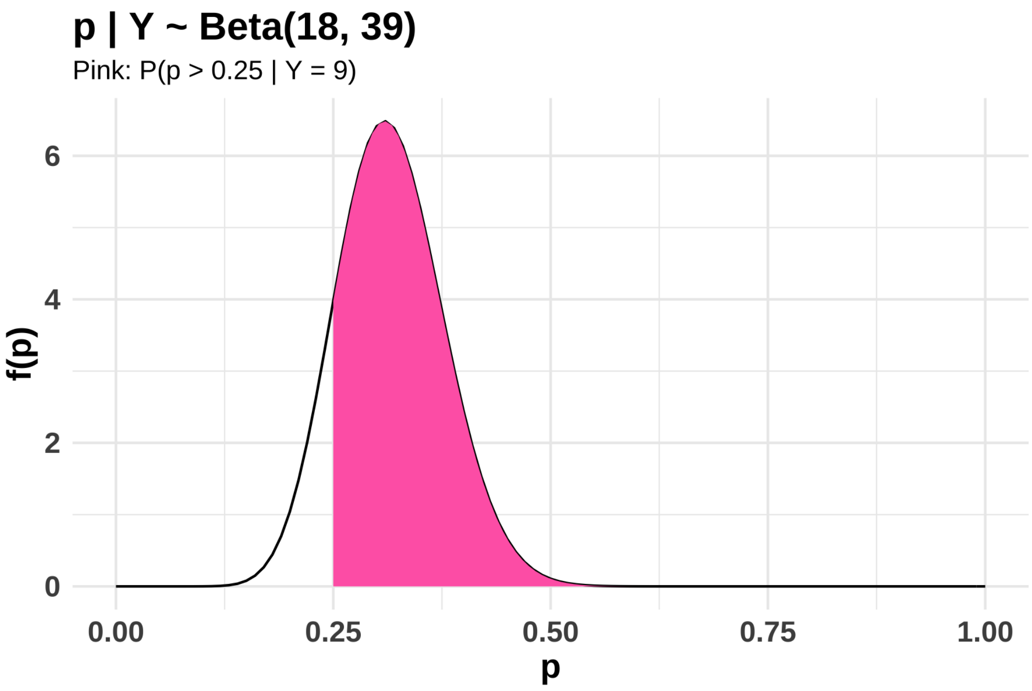



```r
# 95% Credible Interval
qbeta(c(0.025, 0.975), 18, 39)
```

```
## [1] 0.2028545 0.4409674
```

Posterior Probability

While the **visual** of the posterior PDF and the 95% **credible interval** help, we can calculate the **posterior probability** of p being *over 0.25*:




```
1 - pbeta(0.25, 18, 39)
```

```
## [1] 0.8593457
```

Bayesian Hypothesis Testing

We can frame our analysis as two competing **hypotheses**:

- **Null Hypothesis**, $H_0 : p \leq 0.25$ (p is *at most* 25%)
- **Alternative Hypothesis**, $H_A : p > 0.25$

We just calculated the posterior probability of the *alternative hypothesis*:

$$P(H_A \mid Y = 9) = 0.859$$

- Therefore, the posterior probability of the *null hypothesis* is just the **complement**!

$$P(H_0 \mid Y = 9) = 1 - 0.859 = 0.141$$

Posterior and Prior Odds

We can put these two probabilities together to form the **posterior odds**:

$$\text{posterior odds} = \frac{P(H_A \mid Y=9)}{P(H_0 \mid Y=9)} = \frac{0.859}{0.141} = \boxed{6.092}$$

- Our *posterior assessment* is that p is 6.092 times more likely to be *above 0.25* than to be *below 0.25*.

What about the odds **prior** to observing data?? Let's calculate **prior odds**.

- **Recall**: Our *prior* for p was $p \sim Beta(9, 18)$. Then $P(H_A) = P(p > 0.25)$:

```r
1 - pbeta(0.25, 9, 18)
```

```
## [1] 0.8195483
```

$$\text{prior odds} = \frac{P(H_A)}{P(H_0)} = \frac{0.820}{1 - 0.820} = \boxed{4.556}$$

Bayes Factors (BF)

By comparing the *posterior odds* to the *prior odds*, we can obtain the **Bayes Factor**:

$$BF = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_A \mid Y)/P(H_0 \mid Y)}{P(H_A)/P(H_0)}$$

- The Bayes Factor will provide some insight into how much our understanding of R familiarity at Simmons evolved after observing sample data.

```
# Bayes Factor
6.092/4.556 # posterior_odds / prior_odds
```

```
## [1] 1.337138
```

Because the BF is a *ratio*, we should compare it to **1**.

Bayes Factor Scenarios

```r
# Bayes Factor
6.092/4.556 # posterior_odds / prior_odds
```

```
## [1] 1.337138
```

From Bayes Rules!:

1. BF $= 1$: The plausibility of H_A *didn't* change in light of the observed data.
2. BF > 1: The plausibility of H_A *increased* in light of the observed data.
   - The greater the Bayes Factor, the more convincing the evidence for H_A.
3. BF < 1: The plausibility of H_A *decreased* in light of the observed data.

The **posterior probability** of H_A, $P(H_A \mid Y = 9) = 0.859$, is quite high, and the BF > 1 established that the plausibility of my colleague's claim has *increased* in light of the observed data.

⛔BF Cut-Offs ⛔

While the statistics community generally advocates *against* using hypothesis testing to make **rigid conclusions**, there are tables out there that provide cut-off values for BFs.

For example, from Jeffreys (1998):

- BF < 1: Null hypothesis supported😢
- $1 <$ BF $< 10^{1/2}$: Evidence against H_0 "not worth more than a bare mention"🔥
- $10^{1/2} <$ BF $< 10^1$: Evidence against H_0 "substantial"🔥🔥
- $10^1 <$ BF $< 10^{3/2}$: Evidence against H_0 "strong"🔥🔥🔥
- $10^{3/2} <$ BF $< 10^2$: Evidence against H_0 "very strong"🔥🔥🔥🔥
- BF $> 10^2$: Evidence against H_0 "decisive"🔥🔥🔥🔥🔥

Don't actually use cut-offs, though! Science is more *nuanced*, and each hypothesis test and associated *context* should be considered individually. 🔥🔥🔥🔥🔥🔥

⛔Be careful with BFs!⛔

Bayes Factors themselves aren't bad *per se*.

- But **null hypothesis significance testing** can be, and Bayes Factors are often used in this context!

Whether we use the *frequentist* or *Bayesian* perspective, we'll usually be able to reject a "no effect" null hypothesis by **gathering enough data**.

Example

Let's repeat the previous exercise with:

(i) $n = 1000$ and $Y = 300$

(ii) $n = 6$ and $Y = 2$

(In either case, the sample proportion is ~0.3...)

Example

(**Bayes Rules! Exercise 8.9**)

For parameter p suppose you have a Beta(1,0.8) prior model and a Beta(4,3) posterior. You wish to test the null hypothesis that $p \leq 0.4$ versus the alternative that $p > 0.4$.

[*Moose pic to fill space* 👇]

Two-Sided Tests

Two-Sided Hypotheses

Let's revisit the R familiarity example from earlier. But suppose now that we're interested in testing *whether or not* 50% of the students at Simmons have heard of R.

- In other words, we're testing a **two-sided hypothesis**:

$$H_0 : p = 0.5 \quad \text{versus} \quad H_A : p \neq 0.5$$

We've seen this type of test under the **frequentist perspective**, but we hit a roadblock when trying to apply *Bayesian* techniques...

While the Bayesian perspective allows us to calculate the *probability* that H_A or H_0 is true, we can't really do this when p is **continuous**!

😩😩😩

Bayes Factors (BF)

Recall that, by comparing the *posterior odds* to the *prior odds*, we can obtain the **Bayes Factor**:

$$BF = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_A \mid Y)/P(H_0 \mid Y)}{P(H_A)/P(H_0)}$$

However, when we try to calculate the **posterior probability** that H_0 is true, $P(H_0 \mid Y = 9)$, we get zero!

$$P(p = 0.5 \mid Y = 9) = \int_{0.5}^{0.5} f(p \mid y = 9)\, dp = 0$$

- Therefore, the posterior odds are

$$\frac{P(H_A \mid Y)}{P(H_0 \mid Y)} = \frac{1}{0} = \text{🤯🤯🤯}$$

Using Credible Intervals

While we *clearly* can't divide by zero, we could use **credible intervals** to devise an approach to handling two-sided Bayesian hypothesis tests.

Recall that a **95% posterior credible interval** for p is $(0.203, 0.441)$.

```
# 95% Credible Interval
qbeta(c(0.025, 0.975), 18, 39)
```

```
## [1] 0.2028545 0.4409674
```

- Using this, we have a *small* degree of evidence that the H_A is true.
  - The hypothesized value of p, 0.5, falls *just outside* the credible interval.

While one might argue that 0.5 isn't *substantially* outside of the credible interval, we should define what "substantial" means ahead of time!

Bayesian Testing, with a Buffer

Rather than testing

$$H_0 : p = 0.5 \quad \text{versus} \quad H_A : p \neq 0.5,$$

what if instead we tested

$$H_0 : p \in (0.45, 0.55) \text{versus} \quad H_A : p \notin (0.45, 0.55)$$

With this **buffer** in place, we can more rigorously claim *uncertainty* in the plausibility of H_A.

- The hypothesized range of p, $(0.45, 0.55)$, lies *entirely above* the credible interval, $(0.203, 0.441)$.

- We could also calculate the Bayes Factor, since we'll no longer be dividing by zero!