Chapter 7.5 Bayesian Inferences with Continuous Priors

Jim Albert and Monika Hu

Chapter 7 Learning About a Binomial Probability

Introduction

- All Bayesian inferences about the proportion p are based on various summaries of this posterior beta distribution
- We will focus on three types of inference
  - Bayesian hypothesis testing: assess the likelihood of some values of p
  - Bayesian credible interval: find an interval that is likely to contain p
  - Bayesian prediction: learn about new observation(s) in the future.
- The use of simualtion

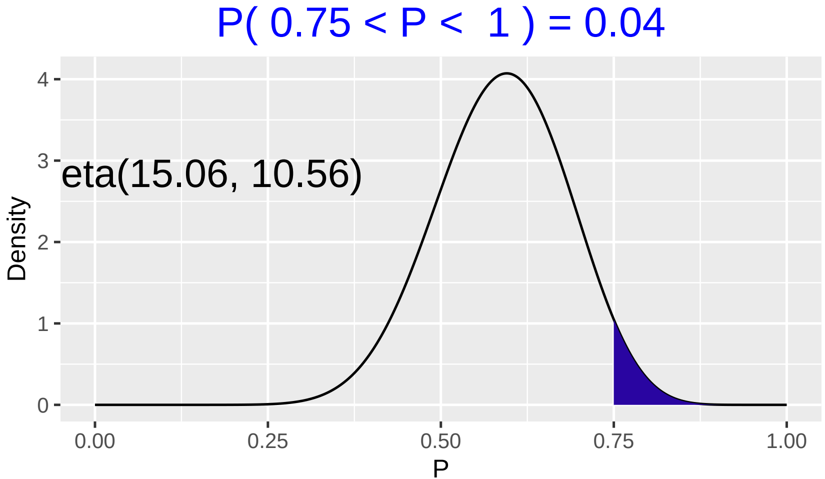
Bayesian hypothesis testing

- Suppose one of the restaurant workers claims that at least 75% of the students prefer to eat out on Friday. Is this a reasonable claim?
- Test the hypothesis $H : p \geq 0.75$
- Bayesian viewpoint: find the posterior probability that $p \geq 0.75$ and make a decision based on the probability

Bayesian hypothesis testing cont'd

- The exact solution: the `beta_area()` function

```
beta_area(lo = 0.75, hi = 1.0,
          shape_par = c(15.06, 10.56), Color = crcblue)
```



Bayesian hypothesis testing cont'd

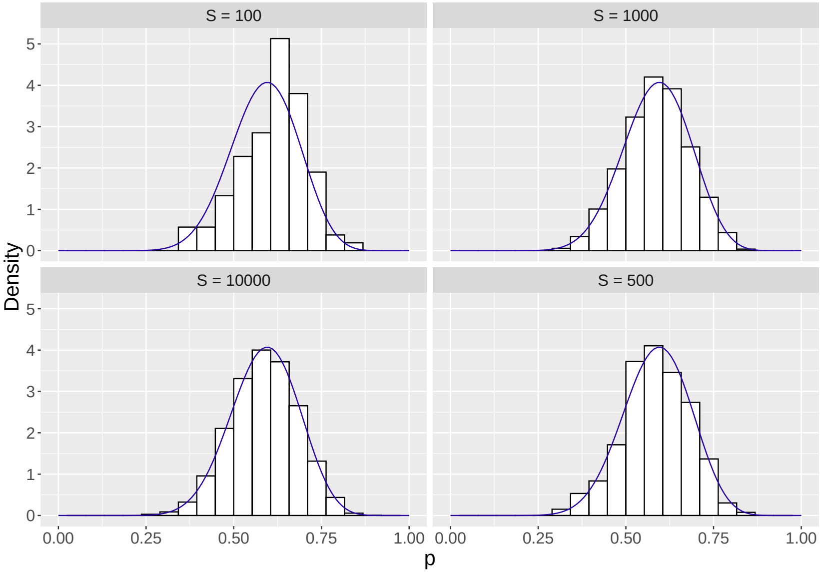
- The simulation solution:
  - generate a large number of random values from the beta posterior
  - summarize the sample of simulated draws to obtain the probability of $p \geq 0.75$

```r
S <- 1000
BetaSamples <- rbeta(S, 15.06, 10.56)
```

```r
sum(BetaSamples >= 0.75) / S
```

```
## [1] 0.051
```

Choice of S



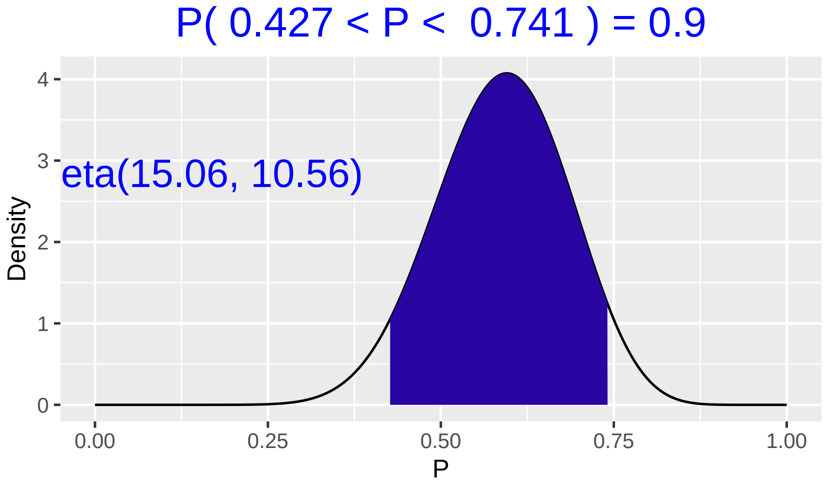
Bayesian credible interval

- An interval provides an uncertainty estimate for the parameter p
- A 90% Bayesian credible interval is an interval that contains 90% of the posterior probability
- Different from the interpretation of a traditional confidence interval

Bayesian credible interval cont'd

- A 90% "equal tails" interval: the `beta_interval()` function

```
beta_interval(0.9, c(15.06, 10.56), Color = crcblue)
```



Bayesian credible interval cont'd

- A 90% "equal tails" interval: the `qbeta()`function

```r
qbeta(c(0.05, 0.95), 15.06, 10.56)
```

```
## [1] 0.4266788 0.7410141
```

Bayesian credible interval cont'd

- A 90% "equal tails" interval: simulation using the `quantile()` function

```
S <- 1000
BetaSamples <- rbeta(S, 15.06, 10.56)
quantile(BetaSamples, c(0.05, 0.95))
```

```
##        5%       95%
## 0.4276337 0.7445882
```

- Choice of S

Bayesian prediction

- Random variable $\tilde{Y}$: he number of students preferring Friday to dine out out of the m respondents
- $\tilde{Y} \mid p \sim$ Binomial(m, p), where p is the **posterior**
- Mathematically,

$$f(\tilde{Y} = \tilde{y}, p \mid Y = y) = f(\tilde{Y} = \tilde{y} \mid p)\pi(p \mid Y = y)$$

$$f(\tilde{Y} = \tilde{y} \mid Y = y) = \int f(\tilde{Y} = \tilde{y} \mid p)\pi(p \mid Y = y)dp$$

- The density of $\tilde{Y}$ given p is Binomial with m trials and success probability p,
- The posterior density of p is $\mathrm{Beta}(a + y, b + n - y)$

Bayesian prediction cont'd

$$f(\tilde{Y} = \tilde{y} \mid Y = y) = \binom{m}{\tilde{y}} \frac{B(a + y + \tilde{y}, b + n - y + m - \tilde{y})}{B(a + y, b + n - y)}$$

- This is the beta-binomial distribution with parameters m, $a + y$ and $b + n - y$

$$\tilde{Y} \mid Y = y \sim \text{Beta-Binomial}(m, a + y, b + n - y).$$

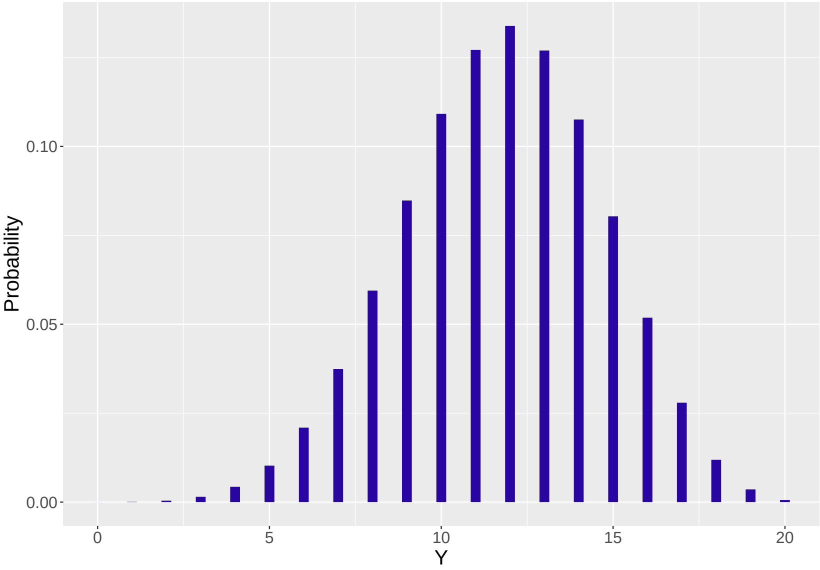
- Summary: Bayesian prediction of a new observation is a beta-binomial distribution where m is the number of trials in the new sample, a and b are shape parameters from the Beta prior, and y and n are quantities from the data/likelihood

Bayesian prediction cont'd

- Compute the predictive probability that $\tilde{y}$ students prefer Friday in a new survey of 20 students
- The exact solution
- The `pbetap()` function from the `ProbBayes` package. The inputs to `pbetap()` are the vector of Beta shape parameters (a, b), the sample size 20, and the values of $\tilde{y}$ of interest.

```
a <- 15.06
b <- 10.56
prob <- pbetap(c(a, b), 20, 0:20)
pred_distribution <- data.frame(Y = 0:20,
                                Probability = prob)
prob_plot(pred_distribution,
          Color = crcblue, Size = 4) +
  theme(text=element_text(size=18))
```

Bayesian prediction cont'd



Bayesian prediction cont'd

- The simulation solution

$$\text{sample } p \sim \mathrm{Beta}(a+y, b+n-y)$$

$$\downarrow$$

$$\text{sample } \tilde{Y} \sim \mathrm{Binomial}(m, p)$$

Bayesian prediction cont'd

- The `rbeta()` and `rbinom()` functions

```
a <- 3.06; b <- 2.56
n <- 20; y <- 12
pred_p_sim <- rbeta(1, a + y, b + n - y)
(pred_y_sim <- rbinom(1, n, pred_p_sim))
```

```
## [1] 10
```

- Repeat this for S times

Bayesian prediction cont'd

- `pred_p_sim` contains 1000 simulated draws from the posterior
- For each element of this posterior sample, the `rbinom()` function is used to simulate a corresponding value of $\tilde{Y}$ from the binomial sampling density

```
a <- 3.06; b <- 2.56
n <- 20; y <- 12
S <- 1000
pred_p_sim <- rbeta(S, a + y, b + n - y)
pred_y_sim <- rbinom(S, n, pred_p_sim)
```

Bayesian prediction cont'd

- Comparison