05 - Binomial analysis

HCI/PSYCH 522
Iowa State University

February 10, 2022

Overview

- Inference for success/failure counts
  - Estimating 1 probability of success
  - Comparing 2 probabilities of success
  - Comparing 3+ probabilities of success

Unknown probability

Suppose you run a study where

- you have n attempts,
- each trial is independent,
- each trial has probability of success θ,

and you are interested in θ.

Let Y be the number of success observed in n attempts and assume $Y \sim Bin(n, \theta)$. A common point estimate is

$$\hat{\theta} = y/n$$

where y is the observed number of successes.

Examples

Suppose you run a study to see how many students correctly register for class using the new Workday system. Since the probability of success might differ depending on what classes need to be registered, you give each student the same list of classes.

- You randomly sample 10 ISU undergraduate students at 8 are successful. Our estimate of the probability of success is $\hat{\theta} = 8/10 = 0.8$.
- You randomly sample 100 ISU undergraduate students at 80 are successful. Our estimate of the probability of success is $\hat{\theta} = 80/100 = 0.8$.
- You randomly sample 1000 ISU undergraduate students at 800 are successful. Our estimate of the probability of success is $\hat{\theta} = 800/1000 = 0.8$.

Although the point estimate is the same, clearly we should have more certainty about the last estimate compared to the first. We need some way to quantify our uncertainty about the true value θ.

Bayesian estimation

Bayes' Rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta)$$

where

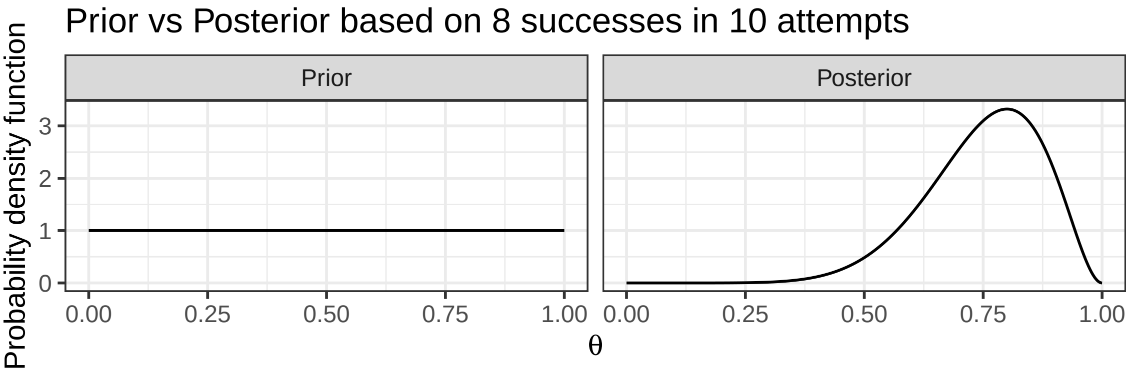
- y is our data,
- θ are our unknowns, e.g. probability of success,
- $p(y|\theta)$ comes from our model, e.g. binomial, (sometimes referred to as the likelihood),
- $p(\theta)$ is our prior belief, and
- $p(\theta|y)$ is our posterior belief.

Thus Bayesian estimation provides a mathematical mechanism to learn about the world using data, e.g.

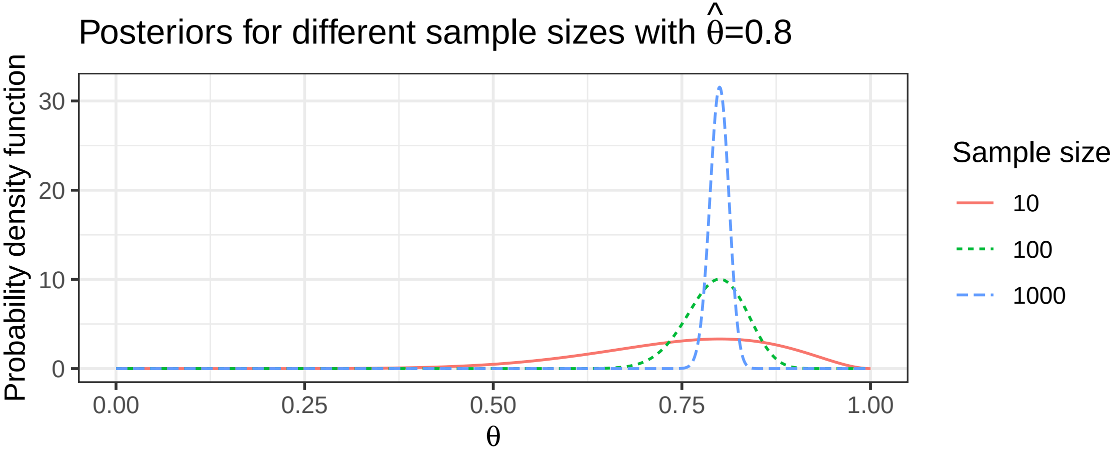
$$p(\theta) \longrightarrow p(\theta|y).$$

Bayesian estimation for probability of success

If we know nothing about our probability of success θ, our prior belief is reasonably represented by a uniform distribution between 0 and 1, i.e. $\theta \sim Unif(0,1)$. When we obtain data y, then our posterior belief is represented by a Beta distribution, i.e. $\theta|y \sim Be(1+y, 1+n-y)$.



Comparison of posteriors



Posterior beliefs

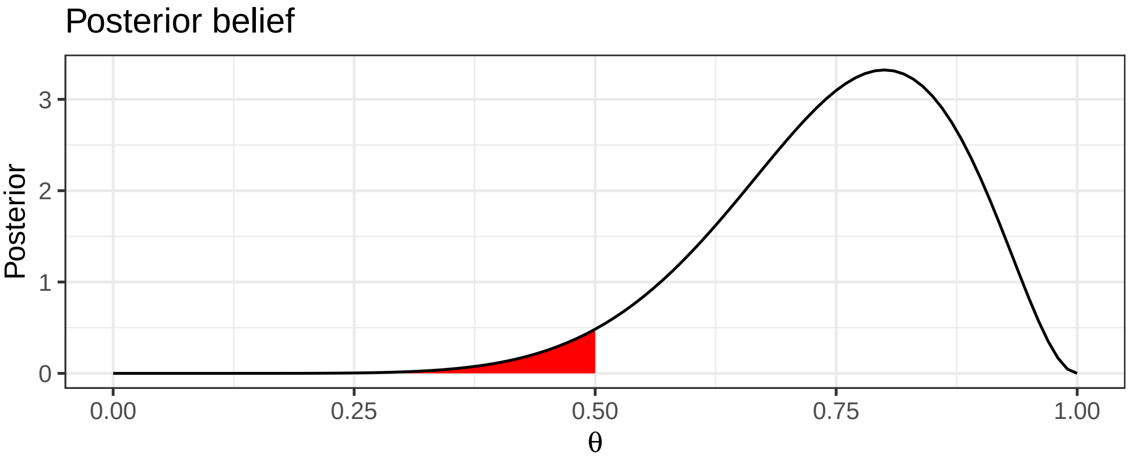
Calculate $P(\theta < c|y)$ for some value c. Let $\theta|y \sim Be(1+8, 1+10-8)$. Calculate $P(\theta < 0.5|y)$:

```
y <- 8; n <- 10; c <- 0.5

pbeta(c, 1+y, 1+n-y)

## [1] 0.03271484
```

Calculate $P(\theta \geq c|y) = 1 - P(\theta < c|y)$.

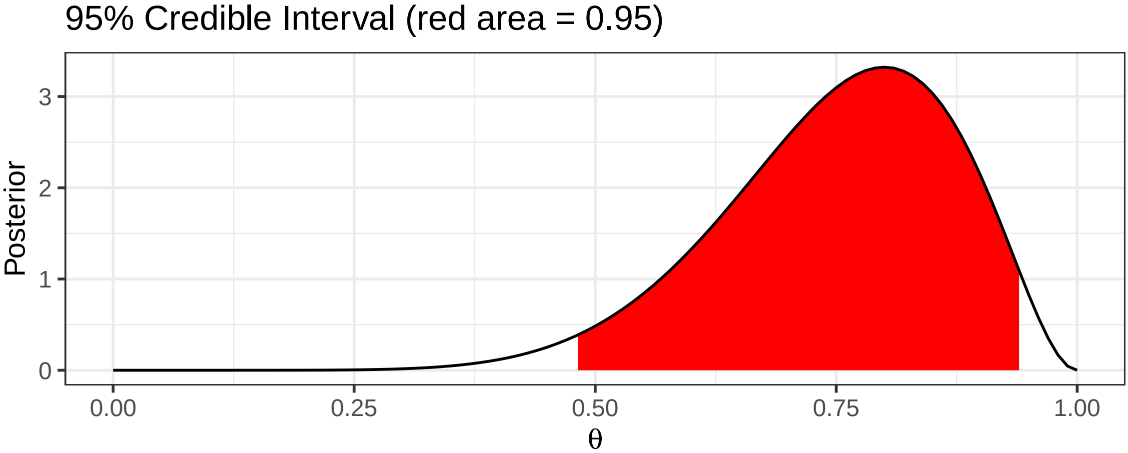
```
1-pbeta(c, 1+y, 1+n-y)

## [1] 0.9672852
```

Posterior beliefs (in a picture)



Credible intervals

A 95% credible interval for θ is the interval such that the area under the posterior is 0.95.



95% Credible Intervals in R

```
a <- 1 - 0.95 # for 95\% CIs

y <- 8; n <- 10
qbeta(c(a/2, 1-a/2), shape1 = 1+y, shape2 = 1+n-y)

## [1] 0.4822441 0.9397823

y <- 80; n <- 100
qbeta(c(a/2, 1-a/2), shape1 = 1+y, shape2 = 1+n-y) %>% round(2)

## [1] 0.71 0.87

y <- 800; n <- 1000
qbeta(c(a/2, 1-a/2), shape1 = 1+y, shape2 = 1+n-y) %>% round(2)

## [1] 0.77 0.82
```

Multiple probabilities

If we are collecting success/failure data under multiple conditions, then we can estimate multiple probabilities.

Let Y_i be the success count in condition i out of n_i attempts for conditions $i = 1, \ldots, \mathrm{I}$. If we assume

- all observations are independent and
- the probability of success within a condition is constant,

then our model is

$$Y_i \stackrel{ind}{\sim} Bin(n_i, \theta_i).$$

If we assume ignorance about θ_i, then we have

$$\text{Prior: } \theta_i \stackrel{ind}{\sim} Unif(0,1) \qquad \longrightarrow \qquad \text{Posterior: } \theta_i | y_i \stackrel{ind}{\sim} Be(1 + y_i, 1 + n_i - y_i).$$

Example

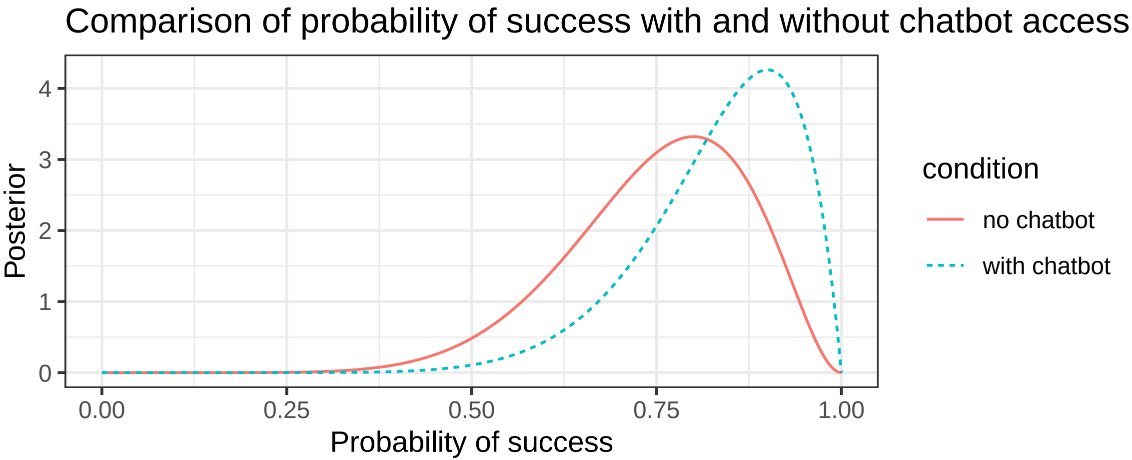
Consider the Workday registration example where we have two conditions:

- no chatbot help and
- with chatbot help.

Research question: How does the chatbot help affect the probability of success in registering for classes?

We randomly select 20 undergraduate students and randomly assign each one a chatbot or no chatbot help condition such that each condition has 10 students (balanced). When we collect the data, we find that 8/10 successfully register without access to chatbot help and 9/10 successfully register with access to chatbot help.

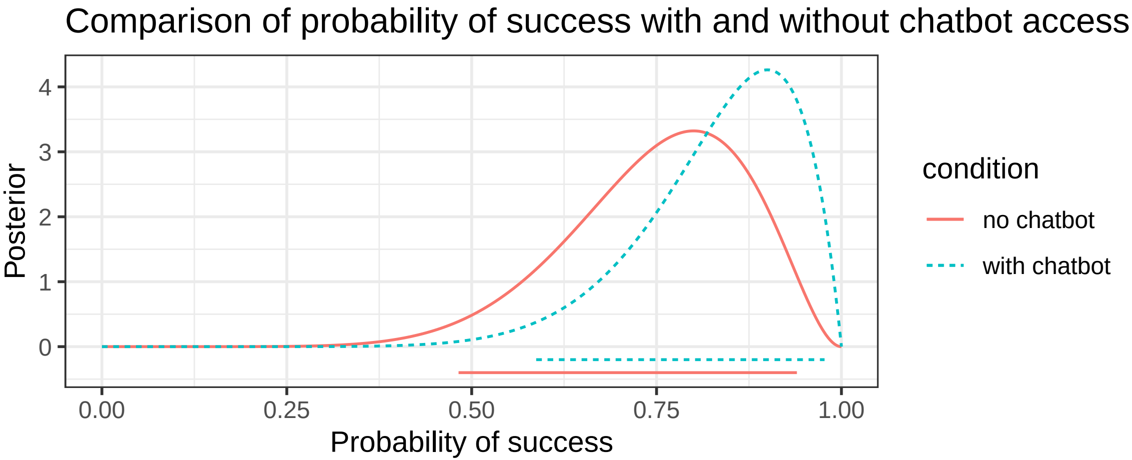
Posterior distributions



95% Credible intervals

```
a = 1-0.95

# no chatbot access
y <- 8
n <- 10
qbeta(c(a/2, 1-a/2), 1+y, 1+n-y) %>% round(2)

## [1] 0.48 0.94

# with chatbot access
y <- 9
n <- 10
qbeta(c(a/2, 1-a/2), 1+y, 1+n-y) %>% round(2)

## [1] 0.59 0.98
```

Plotting credible intervals



Comparing probabilities

Suppose we are interested in calculate

$$P(\theta_{\text{with chatbot}} > \theta_{\text{no chatbot}}|y)$$

where y generally means "all the data".

We can use a Monte Carlo (or simulation) approach:

```
n_reps = 1e5 # some large number
theta_nochatbot   <- rbeta(n_reps, 1+8, 1+10-8)
theta_withchatbot <- rbeta(n_reps, 1+9, 1+10-9)
mean(theta_withchatbot > theta_nochatbot)

## [1] 0.70493
```

How different are the success probabilities?

Rather than just simply knowing if one success probability is larger than the other, we may be interested in knowing how much bigger it is.

We can use the same Monte Carlo samples, calculate the difference, and take quantiles of the result. A 95% credible interval for $\theta_{\text{with chatbot}} - \theta_{\text{no chatbot}}$ is

```
quantile(theta_withchatbot - theta_nochatbot, probs = c(a/2,1-a/2))

##       2.5%      97.5% 
## -0.2331155  0.3985551
```

More than 2 probabilities

Suppose we add the condition of comparing the current registration (through Accessplus?) to the two Workday registration systems (with and without chatbot help).

Research question: How does the Accessplus registration accuracy compare to the two Workday registration options?

Suppose we observe 5/10 successes (with randomly sampled undergraduate students) in the current Accessplus system.

Posterior distributions